UNIVERSITY OF NAIROBI

FACULTY OF SCIENCE & TECHNOLOGY

DEPARTMENT OF MATHEMATICS

STOCHASTIC MODELLING OF SYSTEMATIC MORTALITY RISK UNDER COLLATERAL DATA AND ITS APPLICATIONS

Joab Onyango Odhiambo

August, 2022

A Doctoral Thesis Submitted in the Fulfillment of the Requirements for the Award of the Degree of Doctor of Philosophy in Actuarial Science in the Department of Mathematics, University of Nairobi.
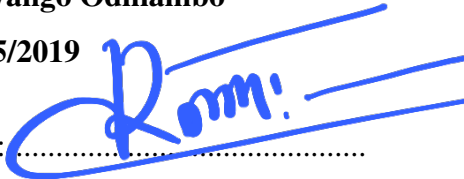
Declaration and Approval

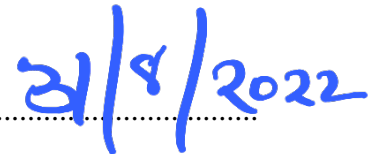I declare that this is my original work and has not been submitted for examination in any other university.

Joab Onyango Odhiambo
I80/55095/2019

Signature:.. Date ....31/8/2022....

Approval

The thesis has been submitted for examination with our approvals as the University Supervisors.
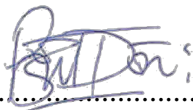
Prof. Patrick G.O. Weke

Department of Mathematics,

University of Nairobi,

Nairobi, Kenya.

Signature .... Date: ....31.08.2022....

Prof. Philip Ngare

Department of Mathematics,

University of Nairobi,

Nairobi, Kenya.

Signature .... Date ....6/9/22....

i

Acknowledgment

Abstract

Many actuaries worldwide use Systematic Mortality Risk (SMR) to value actuarial products such as annuities and assurances sold to policyholders. Data availability plays an essential role in ascertaining the SMR models' accuracy, and it varies from one country to another. Incorrect stochastic modeling of SMR models due to paucity of data has been a problem for many Sub-Saharan African countries such as Kenya, thus prompting modifications of the classical SMR models used in those countries with limited data availability. This study aimed at modelling SMR stochastically under the collateral data environment such as Sub-Saharan African countries like Kenya and then apply it in the current actuarial valuations. This thesis has formulated novel stochastic mortality risk models under the collateral data setup. Kenya population data is preferably integrated into the commonly applied stochastic mortality risk models under a 3-factor unitary framework of age-time-cohort. After testing SMR models on the Kenyan data to assess their behaviours, we incorporate the Bühlmann Credibility Approach with random coefficients in modeling. The randomness of the classical SMR models wass modeled as NIG distribution instead of Normal distribution due to data paucity in Kenya (use of collateral data environment). The Deep Neural Network (DNN) technique solved data paucity during the SMR model fitting and forecasting. The forecasting performances of the SMR models were done under DNN and, compared with those from conventional models, show powerful empirical illustrations in their precision levels. Numerical results showed that SMR models become more accurate under collateral data after incorporating the BCA with NIG assumptions. The Actuarial valuation of annuities and assurances using the new SMR offered much more accurate valuations when compared to those under classical models. The study's findings should help regulators such as IRA and RBA make policy documents that protect all stakeholders in Kenya's insurance, social protection firms, and pension sectors. For areas for further research, one can use the BCA approach for Sub-Saharan African countries with similar demographic characteristics and Hierarchical BCA in SMR modeling.

Table of Contents

List of Tables

List of Figures

Abbreviations and Acronyms

AI: Artificial Intelligence

AIC: Akaike Information Criterion

AIC(c): Modified Akaike Information Criterion

AD: Anderson-Darling

ANN: Artificial Neural Network

AR: Autoregressive Model

ARIMA: Autoregressive Integrated Moving Average

ATC: Age-Time-Cohort

BIC: Bayesian Information Criterion

BCA: Bühlmann Credibility Approach

BC: Bühlmann Credibility

CBR: Central Birth Rates

CDF: Cumulative Distribution Function

CDR: Central Death Rates

CFM: Constant force of mortality

CMI: Continuous Mortality Investigation

CMR: Combined Mortality Rates

DB: Defined Benefit Pension Plan

DC: Defined Contribution Pension Plan

DH: Doornik-Hansen

DL: Deep Learning

DNN: Deep Neural Network

e.g.: for the sake of example

EPV: Expected Present Value

ELA: Endowment Life Assurance

EW: Expanding the Window

HMD: Human Motality Database

i.e.: that is (to say)

i.i.d.: Independent and identically distributed

IRA: Insurance Regulatory Authority

IG: Inverse Gaussian

IMF: International Monetary Fund

JB: Jarque–Bera

LR: Likelihood Ratio

LRM: Linear Regression Mortality Model

LSTM: Long Short-Term Memory

M.L.: Maximum Likelihood

M.L.E.: Maximum Likelihood Estimation

MA: Moving Average Model

MAE: Mean Absolute Error

MAPE: Mean Absolute Percentage Error

MAPFE: Mean Absolute Percentage Forecast Error

ML: Machine Learning

MLP: Multilayer Perceptron

MW: Moving the Window

NIG: Normal Inverse Gaussian

OLS: Ordinary Least Squares

PDF: Probability Distribution Function

RBA: Retirement Benefit Authority

ReLU: Rectified Linear Unit

RMSE: Root Mean Square Error

RNN: Recurrent Neural Networks

RR: Reduction Ratio

SLR: Systematic Longevity Risk

SMR: Systematic Mortality Risk

SVD: Singular Value Decomposition

UDB: Uniform Distribution of Births

UDD: Uniform Distribution of Deaths

VaR: Value at Risk

WLA: Whole Life Assurance

XOR: Exclusive OR

LIST OF PUBLICATIONS

.

Paper I. **Joab Odhiambo,** Philip Ngare & Patrick Weke (2022) Bühlmann credibility approach to systematic mortality risk modeling for sub-Saharan Africa populations (Kenya), Research in Mathematics, 9:1, 2023979, DOI: 10.1080/27658449.2021.2023979.

Paper II. **Odhiambo, Joab**, Patrick Weke, and Philip Ngare. 2021. A Deep Learning Integrated Cairns-Blake-Dowd (CBD) Sytematic Mortality Risk Model. Journal of Risk and Financial Management 14: 259. https://doi.org/10.3390/jrfm14060259

Paper III. **Joab Odhiambo**, Patrick Weke, and Philip Ngare, "Systematic Mortality Risk Modeling Under Age-Time-Cohort Structure":- Accepted at International Journal of Statistics and Applied Mathematics.

Chapter 1

Introduction

In this chapter, we have discussed the background and motivation of the study, actuarial notations, definitions and terminologies used, the statement of the study's problem and objectives, and both general and specific objectives. In addition, we have highlighted the significance as well as the overall structure of the thesis document.

1.1 Motivation and Background

At the start of this century, systematic mortality risk modeling has been of particular significance in actuarial valuations, making it a feature in many; British, European, and American actuarial journals, among many more. Many medical inventions in this century have made it possible for people to live longer than expected by a drop in their death rates (reduced SMR) at the older ages. For example, according to WHO reports, in 2018, the mean life expectancy worldwide increased to 72 years from 68 years in 2010. According to UN statistics in the Department of Economics and Social Affairs in 2018, a male adult in Kenya had a mortality rate of 5695 (per 100,000 male adults), which decreased from 9730 in 2004 Raalte (2021).

Many International Organizations such as World Economic Forum and IMF have also been interested in mortality trends since it affects many countries' fiscal and monetary policies, both developed and developing countries Case and Deaton (2017). In SMR, the company has to correctly determine the number of years in which life is anticipated/expected to survive or die to a given age before deciding the monthly pension money payable to the policyholder. When this is done incorrectly, the pension providers are likely to pay more, thus decreasing the chances of survival due to an increased probability of ruin or insolvency rates.

The current mortality tables used in the Kenyan market may not correctly predict the years

an individual retiree may live after attaining a mandatory retirement age of 60. Today, most insurance companies in Kenya borrow tables from more developed countries like the United States of America, Sweden, the United Kingdom, and other countries. They apply scale factors during actuarial product pricing and valuation. Most policyholders in Kenya often outlive their expected death times since they are borrowed from those countries with different demographic characteristics reports Authority (2017). This effect is compounded by decreased systematic mortality risks that cut across all ages and gender.

Many pension and insurance firms, government, and annuity providers often earn periodically payable annuities in terms of pension money that their surplus process has bought, leading to the application of ruin theory methodology in actuarial life valuation defined by Blake and Hunt (2016). Systematic mortality risk, in many cases, is never easy to understand, transfer, or even try to manage; however, a life assurance company can apply financial derivatives as a financial risk hedging.

With most countries experiencing aging populations, mortality risk has been recognized as among the most common actuarial risks, especially in pension and life assurance mathematics, when formulating pension schemes and life products worldwide. Recently, governments, insurance companies, social security firms, and pension providers worldwide have adopted DC schemes to reduce the risks associated with DB schemes as defined by (Mitchell, 2020) to address reducing global SMR, which is essential in life products pricing.

Many actuaries have a long tradition of using collateral data when improving SMR estimates (Jewell, 1975a). Three main approaches used to accomplish this improvement include model life tables, mortality laws, and relational methods. During the confrontation with estimating SMR in small populations or populations where mortality data do not exist for all the age strata, collateral data offers a solution. In these situations, the population data of the sample alone may not be enough to get reasonable estimates of more than one parameter Bozikas and Pitselis (2020). The condition means collateral data can be used when substituting for limited data. However, most SMR studies are concerned only for short periods, and most populations are always open for departures related to non-death or new group member entrance.

The SMR for people of different ages in life-insured cohorts often constitutes potential collateral data to all other ages within a similar cohort. To use this information, one must have an SMR model Party (2015). SMR Models for different populations exist in two forms, namely deterministic models and stochastic models. Each of these types has various advantages and disadvantages during modeling and projection. While deterministic models are simple to use in modeling SMR, they do not consider the market changes, hence the shortcomings as they do not represent the realities in the market (see literature review chapter). Most actuaries today use stochastic models since they present realities in

the markets, thus leading to correct modeling and pricing of actuarial products beneficial for policyholders.

Incorporating data from a standard life table to compensate for the scanty deaths in the extreme ages is essential for direct data sets. Incorporating collateral data from a standard life table can solve the data paucity problem during the modeling of SMR Najafabadi (2010) and Kim and Jeon (2013). It means the higher the similarity degree between the study population (Kenya) and the standard (the U.K.), the higher the benefits of its use during modeling and actuarial products pricing.

Under collateral data, one can use Credibility theory and machine learning techniques as a solution to data paucity problems. Credibility theory is a topic in actuarial science that uses mathematical modeling to make decisions purely based on historical data Jewell (1975b). While actuaries have been using Credibility theory in their lines of duty, two types of credibility approaches are used when calculating expected risk Buhlmann and Gisler (2005).

On the other hand, the Bühlmann credibility approach looks at the different variances experienced across the population to help actuaries decide depending on the different levels of risks it assesses Ralević (2020).

Deep learning offers experience rating systems new dimensions in SMR modeling by considering individual experiences regarding limited population samples collected Odhiambo, Weke, and Ngare (2021). Credibility theory approaches help use data to improve the estimation accuracy of the conventional models applied in SMR modeling by values, which are reasonable to the extent of using the historical data when forecasting the future SMR Hardy and Panjer (1998). These approaches help customize individual policyholder characteristics according to personal needs, thus increasing actuarial product satisfaction among sophisticated customers in the 21st century.

Recently, governments, insurance companies, social security firms, and pension providers, including Kenya, have adopted DC schemes to reduce the risks associated with DB schemes, as demonstrated by Opoku and Hsu (2019). In many cases, this has been established by legislation where both DC and DB plans are meant to offer employees adequate financial means, thus enabling them to retire and keep a particular standard of living lifestyle during retirement.

If actuaries can model SMR properly, it would be easy to make better projections that would enable the companies to save money, which is often lost from inaccurate estimations when using models that do not consider the different realities of developing markets. Ultimately, this would greatly benefit these insurance firms, thus reducing financial losses that are often experienced today in the world's economic recession, especially after the endemic Covid-19 pandemic, as discussed by Blake and Cairns (2021a).

1.2 Actuarial Notations, Definitions, and Terminologies

1.2.1 Actuarial Notations

l_{x+t}**:** The total number of people who live at a particular age $x+t$, where x, $t = 0, 1, 2.....$

$_t p_x$**:** The probability that an individual of age exactly x will survive to age of $(x+t)$ for all x, $t = 0, 1, 2.....$

$_t q_x$**:** The probability that a person of age exactly x will die between age of x and age of $(x+t)$ for all x, $t = 0, 1, 2.....$

d_x**:** The exact number of people who are aged exactly x will die between age of x and age of $(x+1)$ where $x = 1, 2.....$

1.2.2 Definitions and Terminologies

Systematic Mortality Risk: It is known as the risk of dying earlier than expected.

Force of Systematic Mortality: It is an instantaneous death rate.

Back-Testing: Back-testing is testing a predictive model on specific historical data.

Bootstrapping: Bootstrapping uses random sampling to replace accuracy measures (an error, variance, prediction error, and confidence intervals, among others) with the sample estimate.

Model Robustness: Robustness of a model means strength of a statistical model, procedures, and tests, as per the specific statistical analysis conditions a study aspires to achieve.

A DB scheme: A scheme where benefits are dependant upon on the amount of money submitted to the pension firm, last salary, and years worked.

A DC scheme: It is where your employee and employer's contributions are invested, and the returns are used to buy a pension plan with benefits at the age of retirement.

Collateral Data: Data from a standardized mortality table.

Artificial Intelligence: Artificial Intelligence is the simulation of statistical processes under human intelligence by using computer systems machines.

Machine Learning: This is the use of computer algorithms to enhance statistical experience automatically using data.

Deep Learning: This is a class of ML algorithms, which uses artificial neural networks with data representation learning and forecasting.

1.3 Statement of the Problem

Actuarial modeling and pricing of life products such as annuities and assurances sold in Sub-Saharan African countries like Kenya depend on the existing SMR. The accuracy of SMR depends on the models used and the availability of data during the valuation in the respective Sub-Saharan countries.

Correct modeling of SMR by life assurance companies, pension firms, and government agencies determines the prices of these life products sold to customers. During actuarial pricing of products, overestimating Systsmeatic Mortality Risk will lead to higher rates, resulting in higher costs, making them expensive and unappealing to Kenyans.

Conversely, a lower estimation of SMR will lead to underpricing of these products, making it unsustainable for the companies offering to pay benefits, leading to a higher probability of ruin and insolvency in the long run. This problem calls for correct modeling and forecasting of SMR by incorporating modern methods that will help both parties, namely the companies and policyholders, to have a win-win situation for the growth and satisfaction of the insurance and pension industry in the Sub- Saharan African countries like Kenya.

When correct estimations and valuations of life products are done using the Expected Present Value of assurances and annuities, Kenyans will buy correctly priced annuities and assurance products that serves their needs. The pricing and reserving of assurances and annuities products are done using old period-based assumptions; the liabilities underestimation is expected because of reduced mortality levels.

Suppose actuaries can improve their estimation methods for systematic mortality risk. In that case, they can hedge unpredictable financial losses due to poor product development in Kenyan markets while increasing the number of Kenyans purchasing these products.

1.4 Objectives of the study

1.4.1 General Objective

The study's general objective is to model Systematic Mortality Risk Stochastically Under Collateral Data and its Applications in the Valuation of Actuarial products sold in the Kenyan market.

1.4.2 Specific Objectives

The Specific objectives of the study are to:

1. Model Systematic Mortality Risk Stochastically under the three-factor of Age-Time-Cohort Structure for Kenyan population.

2. Determine Systematic Mortality Risk by Incorporation of Bühlmann Credibility Approach into Stochastic Mortality Models for Kenyan population.

3. Forecast Systematic Mortality Risk Under Deep Learning Technique.

4. Determine the Expected Present Value of Assurance and Annuities under the Integrated Bühlmann Credibility Approach for Kenyan population.

1.5 Significance of the Study

By modeling SMR Stochastically under the Age-Time-Cohort Structure for the Kenyan population, insurance companies can adjust their valuations to deal with the actuarial (life) products sold within the Kenyan market. In addition, insurance and pension players develop reasonably priced pension and life products such as annuities and assurances that work for the Kenyan population.

Incorporating the Bühlmann Credibility Approach helps model SMR for the Kenyan population, and policyholders can adjust policies that protect Kenyans from the insurance companies that take advantage of the uninformed citizens. In addition, to ensure high levels of efficiency, the use of Deep Learning techniques when Forecasting SMR is significant to levels of precision, thus helping insurance companies keep reserves. Regulators such as IRA and RBA can check the reserves held by insurance companies to reduce the firm's ultimate probability of ruin for higher survival chances in a competitive market.

Using correct SMR is essential for the insurance companies when they price the EPV of assurance and annuities under the Integrated Bühlmann Credibility Approach for the Kenyan population. This phenomenon will make the specific insurance company prices competitive, especially in Kenya, with over 50 companies competing in a market with less than ten percent insurance coverage.

1.6 The Structure of the Thesis

This doctoral thesis is organized into seven chapters. Chapter 1 introduces the SMR concept and credibility theory approaches and the benefits of applying the concept to mortality modeling. In addition, it outlined the importance of proper estimation and management in today's financial and actuarial world. It enabled us to look at the general and specific objectives of the research study.

Chapter 2 reviews the actuarial literature on SMR modeling, ranging from some primitive deterministic models and stochastic models used for single age and cohort specifications. We also look at the various gaps identified by the previous researchers and how they were fulfilled.

Chapter 3 explores the SMR Modeling Under the Age-Time-Cohort structure (3-factor systematic mortality risk framework) while considering the Kenyan population set-up. In addition, it looked at the backtesting assumptions on how these mortality models can behave on a population with data paucity using popular SMR models in actuarial research areas.

In Chapter 4, we first introduce the mathematical concept of Buhlmann's credibility to model mortality risk and compared it to the model considering the randomness assumptions. It also looked at the Credibility-based approaches to modeling risk while assuming that randomness does not follow a normal distribution but a NIG (Normal Inverse Gaussian) statistical distribution, a heavy-tailed distribution. The concept of heavy-tailed distribution took into consideration such as shocks in the mortality upsurge. We compared the method with the standard model from the calculated MAPE and RMSE values.

In chapter 5, we forecasted SMR determined Under Deep Learning. We used a simple CBD Model to Forecast SMR Under DL technique, which is essential in calculating life assurance products sold within the Kenyan market. We note the difference between our novel model when compared to the classical models.

In chapter 6, we conducted the actuarial valuation of the products from the projected mortality rates, including life annuities and life assurance products, while comparing how different the values compared to conventionally projected mortality rates. Besides, a comparison of the classical models is determined in the Kenyan set up with the UK data to show the levels of accuracy under the CBA incorporation under the different mortality models in both countries.

In chapter 7, we finally gave general and specific conclusions and recommendations from the study while expanding room for further research for actuaries, academics, and actuarial science studies researchers who wish to continue with studies of SMR modeling.

We have references and projected complete life tables for Kenyans in the Appendix section.

Chapter 2

Literature Review

Modeling human mortality has been vital for many years, starting from the 18th century. In many cases during ancient times, it was done using primitive mathematical methods until Gompertz (1825) brought the definition of a force of mortality as an instantaneous rate of death modeled as an increasing function at an exponential rate. Since then, several mortality models have been developed when modeling the death rates of people by actuaries, statisticians, and demographers. Deterministic models have been used, especially when modeling the mortality risk worldwide. These deterministic models always presuppose that the SMR would be constant during the investigation.

According to Cramér and Wold (1935), the duo primitively attempted a mortality projection and longevity risks by fitting a straight line after studying the Swedish population. However, the results were primitively unrealistic since they suggested that death rates converge at a given age, based on statistical convergence theory.

De Moivre (1725), suggested a survival model applied to the study of actuarial science when modeling population and death rates. He developed a simple mortality law based on a rectilinear survival function modelled using a uniform distribution, allowing deaths to be uniform between two distinct ages.

Gompertz (1825) suggested a Geometric Progression that encompasses mortality risk after a given age. He argued on the physiological grounds that intensifying mortality gained similar proportions in an equal age interval, which gave rise to a progressive force of mortality or instantaneous death rate. However, he argued that mortality rates were exponential between ages 20 and 70 with overestimation for over 80 years, which might not be realistic in the modern world, especially with new medical development and innovations.

William Makeham, in his research study Makeham (1860), did extend the Gompertz (1825) model by adding a constant number to it for more modeling precision. He im-

proved Gompertz's law by introducing a constant and exponentially increasing force of mortality component, thus making causes of death from two reasons: chance and natural deterioration. His model gave a good estimation due to lifestyles experienced at the beginning of the 20th century.

The weaknesses of both the Gompertz and Makeham models were improved Hald (1981), emphasizing an individual model. The author demonstrated systematic mortality in different stages of life, namely young, middle, and old age. He suggested that mortality declines exponentially after survival for the first five years. The models' second part illustrated a mortality increase at young-adult ages from different lifestyle habits such as youth exuberance, drug abuse, exposure to diseases, and accidents from excess drinking, careless driving, and other factors affecting mortality to a greater extent.

In the research of Perks (1932), the author did a linear generalization regression (a logistic regression model) of the Gompertz curve that gave an excellent fit to the mortality risks over an adult's lifetime. It was a way of solving the problem or shortcoming of the number of parameters used to estimate the mortality and longevity risks. Heligman and Pollard (1980) provided a curve with a good fit for all ages when dealing with mortality rates. The model calculates the chance that a given individual at a certain age will die within the next birthday from the current age.

In the most recent actuarial literature of stochastic mortality modeling, many researchers have proposed numerous methods to capture different populations' mortality models. Two American demographers in their publication Lee and Carter (1992) did propose modeling as well as mortality risks forecasting of the aggregate population characteristics of the USA through decomposition of the mortality into specific age-period parameters in what is popularly known as a two-factor mortality model.

As a Lee and Carter (1992) model extension, Cairns, Blake, and Dowd (2006) designed a model that works for higher ages under a two-factor model for the mortality rate modeling and forecasting procedures to smooth the precision of the model developed with cohort effects.

As a follow-up, Tsai and Yang (2015a) did an extension to include cohort effect in the Lee-Carter method to ensure that forecasting results were not overestimated during modeling, whereas Plat (2009) had a proposal of a model that can combine the conventional model to characteristics of Cairns, Blake, and Dowd (2006) and a new stochastic approach dimension. Despite many variants and extensions of Lee and Carter (1992), many researchers have been inspired today by looking for ways of improving the model by introducing extra parameters analysis.

(Hyndman and Ullah, 2007) had applied functional data analysis with penalized regression splines within their framework of modeling mortality and fertility rates using op-

erational data. In addition, (Hatzopoulos and Haberman, 2009)proposed an approach to mortality modeling that operates under the Generalized Linear Models framework with parameterization before forecasting the estimated parameters.

On the other hand, Hatzopoulos and Haberman (2011) did an extension of the Hatzopoulos and Haberman (2009) approach by incorporating the cohort effects in a dynamic approach while assessing how the different cohorts behave and affect the mortality risk of the different population samples under a similar study.

In the last couple of years, several studies have been done to make a comparison of mortality models of many countries depending on dataset characteristics. Booth and Tickle (2008) compared the precision of the forecast through the five Lee-Carter extensions method applying data harvested from ten of the first-world countries. Shang, Booth, and Hyndman (2011) had extended the mortality models for higher accuracy levels through comparison by applying ten techniques afterward, incorporating data from a chosen fourteen countries to compute mean life expectancy.

Additionally, Blake and Cairns (2021b) compared fitting and forecasting of different types of typically applied stochastic models and performed performance analysis for the mortality fluctuations of the US, England, and Wales. The models from Hunt and Blake (2021) had various benefits and shortcomings from the dataset, where they were tested to identify the best that works well during estimation and forecasting.

Gaille (2012) had an application of Lee and Carter (1992) model and the Heligman-Pollard models by modeling Swiss people's mortality rates. In addition, the author compared the forecasts on financial future pension liabilities on the sides of actuaries practicing in the industry and scholars who develop the models for usage in their practicality and respective measurement abilities.

Raalte (2021) made comparison of the forecasts attained from the Lee and Carter (1992) model and the extensions within the correct forecasts found from statistical inferences within developed countries. The author calculated the differences for a particular case of the Dutch population in the expected complete life expectations being realized from the Dutch population.

Hatzopoulos and Haberman (2015) proposed the dynamic parametric model doing the framework of GLM modeling as a tool used to analyze the cohort survival function of mortality for Scandinavian countries such as Norway, Sweden, Finland, and Denmark. The cohort effect affects the quality estimation parameters estimated from the mortality model applied.

(Van Berkum, Antonio, and Vellekoop, 2016) study analyzed the multiple structural changes impact when modeling based on a considerable data scale after filling the models on Belgian and Netherlands male data while analyzing how it affects the quality of the model

used during modeling. Maccheroni and Nocito (2017) did a backtesting on the forecasting Lee and Carter (1992) performance and the Cairns, Blake, and Dowd (2006) models on the Italian data to determine how the specific dataset behaves when modeled.

By taking a given divergence approach to Lee and Carter (1992) variants and their extensions, mortality modeling is applied to credibility theory as an alternative approach, especially when dealing with populations with data limitations. Based on credibility theory applied in actuarial practice, which aims at modeling the periodic flow of limited mortality data, especially for exact age, applying information that originates from a specific range of age span Buhlmann and Gisler (2005).

Hardy and Panjer (1998) applied empirical Bayes credibility as a theoretical basis for risk measures linked with mortality risk for life assurance firms. The author applied Bayesian credibility to solve poor forecasting problems of the mortality risk from the models.

Salhi, Thérond, and Tomas (2016) had a proposal of a credibility approach where they reviewed parameters that fit the Makeham mortality curve, a deterministic model for modeling mortality to determine the behavior when forecasting the expected trends for a specific period. Schinzinger, Denuit, and Christiansen (2016) proposed a multivariate credibility model evolutionary for mortality enhancement rates, describing the joint mortality dynamics in different countries' populations.

Hardy and Panjer (1998) proposed a Bayesian non-parametric model used in those smaller populations and multi-dimensional data when modeling mortality rates at the same considering a stable mortality table benchmark of a large and developed population serving as the prior information in the data. By applying an adaptive smoothing approach based solely on a local likelihood, similarly, Salhi and Thérond (2018) had a proposal on the existing methodology of adjusting the graduated mortality table numbers. It was based on the credibility techniques and giving essential highlights on the benefits of applying credibility approaches in pricing life assurance, pension plans, and annuity products for individuals.

Li and Lee (2005) did propose an approach of a two-step procedure when modeling mortality risk dynamics for more than one population by extending the Lee and Carter (1992) to include coherent forecast of mortality risks for different populations. Many contributions are found in actuarial literature, where many researchers have made massive contributions that regard mortality modeling for multi-dimensional populations. Li and Hardy (2011) had an assumption of a linear relationship between the base population of the time-varying index and many other common populations when estimating mortality while developing ways to hedge the SMR.

Hunt and Blake (2021) made a proposal of a Bayesian framework that can mutually model two simultaneous populations while making comparisons of the standard mortality mod-

els commonly applied by actuaries and demographers today. Hatzopoulos and Haberman (2013) did a presentation of a coherent structure of mortality risk modeling and forecasting under the Generalized Linear Model structure for mortality dynamics analysis applying global data from the HMD website and finally made a comparison on the different mortality risk dynamics.

Li, Zhou, and Hardy (2015) did a generalization of a single-population mortality risk model in diverse probable ways to fit two or even more populations before measuring them based on individual risk before applying the hedging of longevity products.

Credibility Mortality modeling offers an opportunity for actuaries to improve their models, thus offering them a chance to value the life assurance products much easier without too many challenges as experienced in the past number of years as illustrated in the book Buhlmann and Gisler (2005) and Jewell (1975b). In addition, Levantesi and Pizzorusso (2019) did an application of machine learning techniques to systematic mortality risk modeling and forecasting.

(Tsai and Lin, 2017a) made an application of the concept of Bühlmann's credibility to mortality data of three countries, namely Japan, the US, and the UK, at the same time making a comparison on the quality of the estimates from the conventional model. Tsai and Lin (2017b) had also published a paper within the same year that integrated Bühlmann credibility approach into the Lee and Carter (1992) model to enhance the precision of the mortality models from the new novel approach, the linear relational model of liner regression approach to credibility theory during mortality modeling.

Cairns, Blake, and Dowd (2006) model is a simple way of enhancing its forecasting performance during prediction for a dataset for the UK, assuming that the randomness of the model follows a Gaussian distribution. Recent actuarial science contributions of modeling mortality risk under a credibility incorporated framework were demonstrated by Tsai and Lin (2017b) and Tsai and Zhang (2019), in Tsai and Lin (2017b) paper, where the authors integrated the Bühlmann credibility approach to mortality data of the three countries, namely Japan, the US, and the UK.

According to the study Tsai and Zhang (2019), the authors did the Bühlmann credibility approach integration into the Lee and Carter (1992) model, the Cairns, Blake, and Dowd (2006) and Tsai and Yang (2015b) models to improve individual predictability when predicting performance and precision for the US dataset.

Based on the previous studies, we have brought in the concept of 3-factor SMR modeling, the BCA approach to SMR modeling, and the deep learning technique in SMR prediction. All these novel methods are aimed at improving the existing modelling for more accurate results than the classical models, meaning that we have better estimates that will make the SMR more precise during actuarial modelling and valuation.

Chapter 3

Systematic Mortality Risk Modeling Under Three-Factor Structure

In this chapter, we introduce the novel modelling concept of Systematic Mortality Risk Modeling Under the three-factor of Structure of Age-Time-Cohort using three popular classical SMR models. We determine the behavior of the classical stochastic Mortality Risk Models under the Kenyan setup and test whether the characteristics still hold despite the data paucity challenges. We define the classical SMR models, fit them in Kenyan data to estimate parameters before testing for Robustness properties, and then use Information Criteria when selecting the most parsimonious model.

We forecast the SMR under Classical models and analyze results showing that paucity of data affects the forecasted behavior of SMR. Results show that SMR fits the Kenyan population setup well compared to a two-factor model.

3.1 Stochastic Mortality Risk Models

This section defines the popular classical stochastic mortality models for modeling SMR;

Definition 3.1.1. Let $\mu(x,t)$ be a stochastic mortality model in which the natural logarithm of a time series, especially of age-specific death rates, equals the sum component of age-specific and a part of a time-varying parameter product Lee and Carter (1992) defined as;

$$\mu(x,t) = \alpha_x + \beta_x k_t + e(x,t) \qquad (3.1.1)$$

where α_x describes the mortality pattern of age-specific group, k_t time-trend of mortality index model levels, β_x shows the increase in mortality of a person aged x and $e(x,t)$ is an error term following a White Noise with mean of $\mu = 0$, with variance of σ^2.

$$x = 1,2,3,.......w \text{ and } t = 1,2,3,........n$$

Since the model does not fully identify its parameters, it necessitates the needs of constraints of $\sum\limits_{x=1}^{\infty} \beta_x = 1$ and $\sum\limits_{t=1}^{\infty} k_t = 0$.

Definition 3.1.2. Let $\mu(x,t)$ be a two-factor stochastic model that uses the logit transformation of mortality risk instead of the natural logarithms in systematic mortality risk modeling Cairns, Blake, and Dowd (2006) defined as;

$$\mu(x,t) = \alpha_x^{(1)} k_t^{(1)} + \alpha_x^{(2)} k_t^{(2)} \tag{3.1.2}$$

where the $\alpha_x^{(1)} = 1$ and $\alpha_x^{(2)} = (x - \bar{x})$ are the parameter assumptions.

The equation (3.1.2) can be rewritten under parameter assumptions to become;

$$\mu(x,t) = k_t^{(1)} + (x - \bar{x}) k_t^{(2)} \tag{3.1.3}$$

It is much easier to estimate the parameters of $\alpha_x^{(2)}$ which is just the $(x - \bar{x})$ as the observed variables minus the expected number of the variables. To consider the effects of cohort in a given population, the model is simplified as;

$$\mu(x,t) = \alpha_x^{(1)} k_t^{(1)} + \alpha_x^{(2)} k_t^{(2)} + \alpha_x^{(3)} w_{t-x}^{(3)}$$

With the above assumptions of parameters taking the forms; $\alpha_x^{(1)} = 1, \alpha_x^{(2)} = (x - \bar{x})$ and $\alpha_x^{(3)} = 1$, the equation becomes;

$$\mu(x,t) = k_t^{(1)} + k_t^{(2)}(x - \bar{x}) + w_{t-x}^{(3)} + e(x,t) \tag{3.1.4}$$

with $w_{t-x}^{(3)}$ as the cohort effect and $e(x,t)$ as the error term.

Definition 3.1.3. Let $\mu(x,t)$ be a stochastic mortality model expressed by Tsai and Yang (2015a) be given as;

$$\mu(x,t) = ln\,(m(x,t)) = \kappa_t^{(0)} + \kappa_t^{(1)} ln\,((m(x,t_x - 1)) + e(x,t) \tag{3.1.5}$$

In mortality modeling, we apply the constraint $\sum\limits_{i=1}^{\infty} w_{t-x}^{(3)} = 0$ to prevent presenting the problem of identifiability during both estimation and forecasting process and where $x = 1,2,3,....n$, and $t = 1,2,3....m$ denotes the base year while the parameters $\kappa_t^{(0)}$ and $\kappa_t^{(1)}$ are obtained as the regression coefficients from $ln(m(x,t))$ on a line of $ln(m(x,t-1))$ for values $t = 1,2,3,.....m$. It satisfies the precondition of the value;

$$\hat{k}_t^{(0)} + \frac{1}{n}\hat{k}_t^{(1)} \sum\limits_{x_1}^{x_n} ln\,(m(x_i,t_x - 1)) = \frac{1}{n} \sum\limits_{x_1}^{x_n} ln\,(m(x_i,t)) = \frac{1}{n} \sum\limits_{i=1}^{n} ln\,(m(x_i,t)) \tag{3.1.6}$$

for $i = 1, 2, \ldots n$.

The following are the popular reviewed stochastic mortality models used in actuarial studies:

Name of the Model	Mathematical Formula	Notation
Lee and Carter (1992)	$\mu(x,t) = \alpha_x + \beta_x k_t + e(x,t)$	**A**
Cairns, Blake, and Dowd (2006)	$\mu(x,t) = k_t^{(1)} + k_t^{(2)}(x - \bar{x}) + e(x,t)$	**B**
Tsai and Yang (2015a)	$\mu(x,t) = \kappa_t^{(0)} + \kappa_t^{(1)} ln(m(x, t_x - 1)) + e(x,t)$	**C**

Table 3.1: Summary of Popular Stochastic Mortality Models

In this study, we use these mortality models tabulated in Table 3.1 since they are among the top models that many actuaries use today when modeling SMR Case and Deaton (2017).

3.2 Force of Systematic Mortality

Proposition 3.2.1. *Let $\mu(x,t)$ be force of systematic mortality risk defined as an instantaneous survival rate for a personal life aged exactly x, then its future lifetime is given by*

$$\mu(x,t) = \frac{{}_t p_x}{f(t)} \qquad \forall \, x, \, t \geq 0$$

where ${}_t p_x$ is the survival probability and $f(t)$ being the future lifetime distribution of SMR.

Proof: The derivation of force of systematic mortality risk is as follows,

$$\mu(x,t) = \lim_{\Delta_t \longrightarrow 0} P\left[\frac{t \leq K \leq t + \Delta_t, \delta = 0 \ /K > t}{\Delta_t}\right] \qquad (3.2.1)$$

$$\mu(x,t) = \lim_{\Delta_t \longrightarrow 0} \frac{1}{\Delta_t} {}_t p_x$$

From the definition of ${}_t p_x$, it is possible to define it in terms of l_x, where the value of ${}_t p_x = \frac{l_{x+t}}{l_x}$ and δ is the force of interest. We rewrite the equation in this way;

$$\mu(x,t) = \lim_{\Delta_t \longrightarrow 0} \frac{1}{\Delta_t}\left[\frac{l_{x+t}}{l_x}\right] \qquad (3.2.2)$$

By introducing $\frac{l_x}{l_x} = 1$ on both sides of the equation (3.2.2) with a negative before rearranging to take the form of a derivative form;

$$\mu(x,t) - 1 = \lim_{\Delta_t \longrightarrow 0} \frac{1}{\Delta_t}\left[\frac{l_{x+t}}{l_x} - 1\right]$$

15

$$\mu(x,t) - 1 = -\frac{1}{l_x} \lim_{\Delta_t \longrightarrow 0} \left[\frac{l_x - l_{x+t}}{\Delta_t} \right] \tag{3.2.3}$$

It is simple to make $\mu(x,t)$ the subject in the above equation (3.2.3) as;

$\mu(x,t) = 1 - \left(\frac{1}{l_{x+t}} \right) \frac{d}{dt}(l_{x+t})$

$$\mu(x,t) = 1 - \left(\frac{l'_{x+t}}{l_{x+t}} \right)$$

$$\mu(x,t) = 1 - \left(\frac{s'(x+t)}{s(x+t)} \right) \tag{3.2.4}$$

where the value of $s'(x+t)$ is the derivative of the survival functions of a life aged $x+t$ and $s(x+t)$ are the survival functions of a life aged $x+t$, the l'_{x+t} is the derivative of the l_{x+t}, which are the number of individuals surviving at a specific age $x+t$. From derivation of force of systematic mortality, it is important to use the formula when calculating the value of $_t p_x$ as;

$$_t p_x = \left(\frac{l_{x+t}}{l_x} \right) = e^{-\int_{\forall s} \mu(x,s)ds} dt$$

Expressing this *PDF* as a future lifetime of x in terms of force of systematic mortality risk as;

$$\mu(x,t) = \begin{cases} \frac{_t p_x}{f(t)}; & x,t \geq 0 \\ 0 & elsewhere \end{cases} \tag{3.2.5}$$

Hence, the proof.

Lemma 3.2.1. *Let e_x^o be the complete expectation of life and be of the form of $\int\limits_0^\infty {}_t p_x dt$ for $x, t \geq 0$. Then $e_x^o = \int_0^\infty e^{-\int_0^t \mu(x,s)ds} dt$*

Proof. By applying the expectation of life rule, we have; $\qquad\qquad\qquad\qquad$ □

$$e_x^o = \int_0^\infty t f(t)dt$$

Substituting equation (3.2.5), we have;

$$e_x^o = \int_0^\infty t \times {}_t p_x \mu(x,t)dt$$

We then integrate the expectation of life equation using integration by parts as;

Let $u = t$ and $dv = {}_t p_x \mu(x,t)$

$du = dt$ and $v = -{}_t p_x$

Using integration by parts, we get;

$$\implies \int_0^\infty t \times {}_t p_x \mu(x,t) dt = [-t \times {}_t p_x]_0^\infty + \int\limits_0^\infty {}_t p_x dt = e_x^o = \int\limits_0^\infty {}_t p_x dt$$

but the probability that one person aged x surviving up to an age of $x+t$ can be written as a summation of the force of systematic mortality over the given period of life.

$$e_x^o = \int_0^\infty e^{-\int_0^t \mu(x,s)ds} dt$$

Hence, the lemma is proved.

Proposition 3.2.2. *Let e_x be the curtate expectation of life and be of the form of $\sum\limits_{t=1}^\infty {}_t p_x$ for all value of $x = 0,1,2,...,\infty$. Then e_x will be $\sum\limits_{t=1}^\infty {}_t p_x = \left(\frac{l_{x+1}+l_{x+2}+l_{x+3}+...}{l_x} \right)$.*

Proof. From the theoretical definition of curtate expectation of life, we have;

$$e_x = 1\left(\frac{(l_{x+1})(q_{x+1})}{l_x} \right) + 2\left(\frac{(l_{x+2})(q_{x+2})}{l_x} \right) + 3\left(\frac{(l_{x+3})(q_{x+3})}{l_x} \right) + ...$$

$$= \frac{1}{lx}\{(l_{x+1})(q_{x+1}) + 2(l_{x+2})(q_{x+2}) + 3(l_{x+3})(q_{x+3}) + ...\}$$

$$= \frac{1}{lx}\{(l_{x+1})(q_{x+1}) + 2(l_{x+2})(q_{x+2}) + 3(l_{x+3})(q_{x+3}) + ...\}$$

$$= \frac{1}{l_x}\left\{ \begin{array}{l} [(l_{x+1})(q_{x+1}) + (l_{x+2})(q_{x+2}) + (l_{x+3})(q_{x+3}) + ...] + \\ [(l_{x+2})(q_{x+2}) + (l_{x+3})(q_{x+3}) + ...] + [(l_{x+3})(q_{x+3}) + ...] \end{array} \right\}$$

$$e_x = \frac{1}{l_x}\{l_{x+1} + l_{x+2} + l_{x+3} + ...\}$$

We obtain the results as follows;

$$e_x = \left(\frac{l_{x+1}+l_{x+2}+l_{x+3}+...}{l_x} \right) \implies e_x = \sum_{t=1}^\infty {}_t p_x.$$

Hence, the proof. $\square$

3.3 The Age-Time-Cohort Modeling Framework of SMR

3.3.1 Introduction

Generally speaking, actuaries today focus on modern model development methods, which could help estimate future SMR trends of different populations. Going in the similar direction of Rutherford, Lambert, and Thompson (2010) and Hunt and Blake (2020), improved the concept of Age-Time-Cohort (ATC) characterization scheme for the prevailing models in the SMR measurement research. This makes the concept of ATC mortality modeling vital since it is a three-factor mortality modeling instead of a one or two-factor modeling procedure in terms of accuracy and simplicity in analyzing the results.

We have organized the rest of this chapter in the following way. Section (3.4) describes fitting these models, whereas Section (3.5) shows some findings of the mortality projection for each classical model. The findings compared with those from the conventional studies are tabulated in Section (3.4). In the end, Section (3.6) shows concluding remarks from the analyzed results based on the Kenyan data.

3.3.2 The Age-Time-Cohort Mortality Modeling

Definition 3.3.1. Let $d(x,t)$ be the total number of observed deaths between age x and t while $\theta(x,t)$ be the central population exposed to these deaths (at the midpoint of year t). We have approximated the initial population exposures through $\theta^0(x,t) \approx \theta(x,t) + \frac{1}{2}d(x,t)$. Thus, we get a single year death probability at age x defined as $q(x,t) = \frac{d(x,t)}{\theta^0(x,t)}$ whereas death rate $m(x,t) = \frac{d(x,t)}{\theta(x,t)}$. Blake and Cairns (2021b) made a UDD assumption and a constant force of systematic mortality risk between any two consecutive integer ages under ATC structure.

A stochastic *ATC* model can resemble a response random variable that is well-defined as a single-year death probability $q(x,t)$ or the consistent force of systematic mortality $\mu(x,t)$ to be the best predicting estimate, which is dependent upon the ages from on age $x = x_1$,$x_2, x_3....x_{n-1}, x_n$, time $t = t_1, t_2, t_3..., t_{m-1}, t_m$, and cohort effects (a person aged x year of birth) $w = t_1 - x_n, t_2 - x_{n-1}, t_3 - x_{n-2}, ..., t_{m-1} - x_2, t_m - x_1$ for a given set of population.

Definition 3.3.2. Let $\lambda(x,t)$ be the *ATC* framework defined as;

$$\lambda(x,t) = \alpha_x + \sum_{j=1}^{K} \beta_x^{(j)} k_t^{(j)} + \beta_x^{(0)} \gamma_{(t-x)} \tag{3.3.1}$$

where $\lambda(x,t)$ is the link function that changes the measure of systematic mortality risk into an appropriate modeling form, α_x is the non-dynamic age function, which is responsible for expressing the general mortality shape by age, $\beta_x^{(j)} k_t^{(j)}$ is the set of K age-time terms, which determines the mortality trends with $k_t^{(j)}$ indicating the general mortality pattern with time whereas $\beta_x^{(j)}$ showing the mortality change patterns across ages.
The function $\beta_x^{(0)} \gamma_{(t-x)}$ is denoted as age-cohort term, whilst $\gamma_{(t-x)} = \gamma_w$ is capturing the effects of every person at a given year of birth w and $\beta_x^{(0)}$ modifies the general cohort effect across all ages.

Several factors need to be considered when selecting the response variable, which the link function of $\lambda(x,t)$ transforms it and depends on the mortality data representation. We can assume that numbers of deaths experienced in a particular population denoted by $D(x,t)$ follows a Binomial distribution random variable at age x to age $(x+t)$ year having

parameters $(\theta^0(x,t), q(x,t))$. From the Binomial distribution of $\mathbb{E}(D(x,t)/\theta^0(x,t)) = q(x,t)$, it is easy to apply the initial exposures of $\theta^0(x,t)$. With the assumption that the random variable $D(x,t)$ follows a Poisson distribution having parameter $(\theta(x,t)\rho(x,t))$, it is easy to obtain the mean (expected value of deaths) as $\mathbb{E}(D(x,t)/\theta(x,t)) = \rho(x,t)$, where the central exposures used is $\theta(x,t)$.

From the assumption of the Binomial distribution and the logit expression used in defining the probability of death as well as link function taking the form of $\lambda(x,t) = $ logit $q(x,t)$ $= log \frac{q(x,t)}{(1-q(x,t))}$, whereas for a Poisson distribution, the total deaths number experienced will be assumed to be $\lambda(x,t) = log \rho(x,t)$. According to Hunt and Blake (2020), we group the models as Generalized non-linear since it has a presence bi-linear term of $\beta_x k_t$ in the model.

3.3.3 Data Integrity and the Assumptions

Let the number of deaths observed during the study be $d(x,t)$ whereas the central exposures $\theta(x,t)$ for the Kenyan dataset. To achieve model consistency during modeling comparisons, all models must possess the same statistical distribution assumptions, and modeling results need to use similar mortality measures proposed by Haberman and Renshaw (2011).

Thus, we assume that total deaths follows Binomial probability distribution using the link function of $\lambda(x,t) = $ logit $q(x,t)$. During this study, we use those ages from 60 and 100 since most Kenyans retire at the age of 60 and live henceforth until death while enjoying the retirement benefits in the form of annuities payable as long as a life aged x is still then alive. We back-test the model using the Kenyan population complete life table data to determine their reliability.

Additionally, Blake and Cairns (2021b) had pointed out that data reliability of all cohort parameters estimated from $\gamma_{(t-x)}$ or γ_w that solely depends on the total observations made for every birth year. Before doing any analysis, we repeated the same procedure while excluding those cohorts with less than five to nine observations since our Kenyan datasets have shorter periods. It excludes more than six cohorts that may be excessive to offer an optimal balance of the Kenyans between the fitting model and forecast its behavior.

While this model choice will provide similar fitting results compared to a case where fewer cohorts are excluded, resulting in reasonable forecasts, which prevent over-fitting the effects of the cohort.

3.3.4 Back-Testing of Age-Time-Cohort Mortality Models

From *Table* 3.1, all deaths in these models are assumed to be following a Binomial statistical distribution with $\lambda(x,t) = logit\, q(x,t)$. For purposes of diagrammatic illustrations, we denote the SMR models using letters as also shown in Table 3.1.

3.4 Mortality Model Fitting

In the ATC model fitting section, we fit our models after estimating the respective parameters of the different mortality models. We estimate model's A parameters using the method of SVD in the OLS context of fitting, whereas model B and C are minimized through the deviance of the respective predictor structure. From the method proposed by Brouhns, Denuit, and Vermunt (2002), we estimate age, time, and cohort parameters through maximizing the likelihood functions of the respective models.

Corollary 3.4.1. *Let $\mathscr{L}(\Theta)$ be the log-likelihood for models from A, B, and C considering the Binomial distribution assumptions as given Villegas, Kaishev, and Millossovich (2015), then;*

$$L(\Theta) = \sum_{\forall x,t} \varpi_{x,t} \left\{ \Theta * ln(\frac{\hat{\Theta}}{\theta^0(x,t)}) + [\theta^0(x,t) - \Theta] * ln(\frac{\theta^0(x,t) - \hat{\Theta}}{\theta^0(x,t)}) + ln \begin{pmatrix} \theta^0(x,t) \\ \Theta \end{pmatrix} \right\}$$

where $d(x,t) = \Theta, \hat{d}(x,t) = \hat{\Theta}$, and $\theta^0(x,t)$ is the initial exposure of the population, whereas $f^{-1}(u)$ denoting the specific inverse link function $f^{-1}(u) = logit\, u$ and u is the ATC structure.

Therefore, the mean deaths for every $(A, B, $ and $C)$ models is given as follows;

$$\hat{\Theta} = \theta^0(x,t)f^{-1}(\lambda(x,t)), \text{ which is replaced to yield}$$

$$\hat{\Theta} = \theta^0(x,t)f^{-1}\left(\alpha_x + \sum_{j=1}^{K} \beta_x^{(j)}k_t^{(j)} + \beta_x^{(0)}\gamma_{(t-x)} \right) \tag{3.4.1}$$

With $K = 1$ for, $N = 2$ for A, C and $K = 2$ for B, and the respective prior weights denoted by $\varpi(x,t)$ and can take a value 0 for the weight of the excluded data-cell and 1 for the inclusive data-cell.

During fitting models A, B, and C, we follow the assumption of Poisson with $\lambda(x,t) = \log \mu(x,t)$, since the models have been adjusted from the conventional studies to get the exact estimates of parameters. Furthermore, using the given periods and shown in Figures (3.4.1–3.4.3) determined robustness of estimates of the used parameters

By obtaining MLEs of the estimates with the Binomial distribution assumption for *A*, *B*, and *C* models respectively, for Kenyan males and females aged 60 years to 100. Upper lines matches to the estimates of parameters for the fitted during, whereas lower lines for the same period. From the diagram, we get the explanatory variations on the parameter estimates.

On the fitting of the data to the model, we start with model *A*, which is fitted as illustrated below as follows:
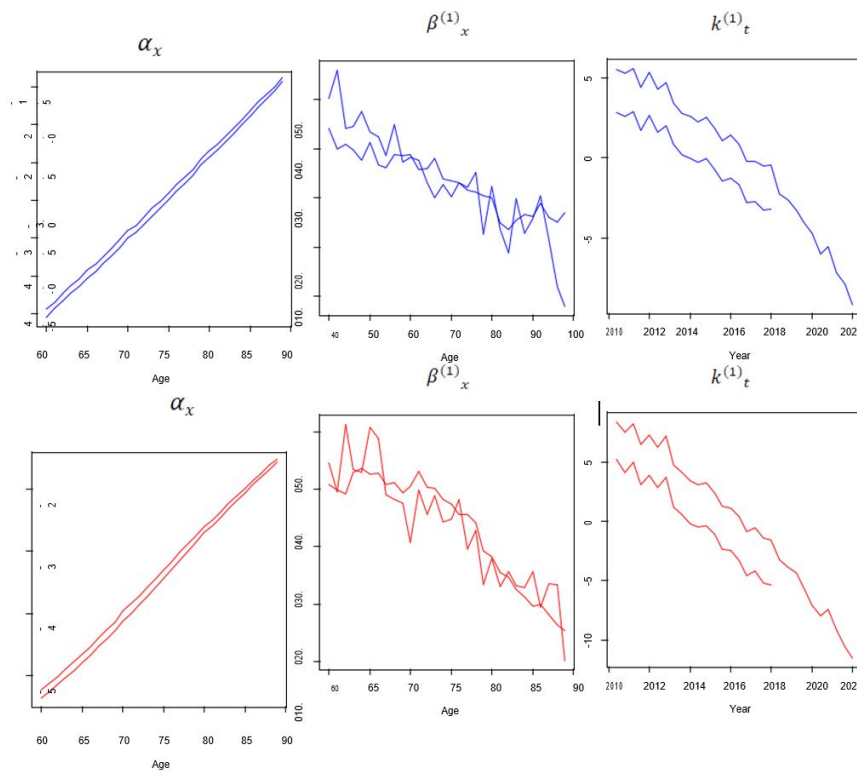
Figure 3.4.1: α_x, $\beta_x^{(1)}$ and $\kappa_t^{(1)}$ parameters estimates for males (blue top panels) & females (red bottom panels) for personal ages fitted from 2010 to 2022 for Model *A*

The estimates of α_x in (Figure 3.4.1) above demonstrates an approximately linear upward trend in mortality for both males and females. Both $\beta_x^{(1)}$ and $\kappa_t^{(1)}$ show a downward trend, which shows the general decline in mortality among both males and females.
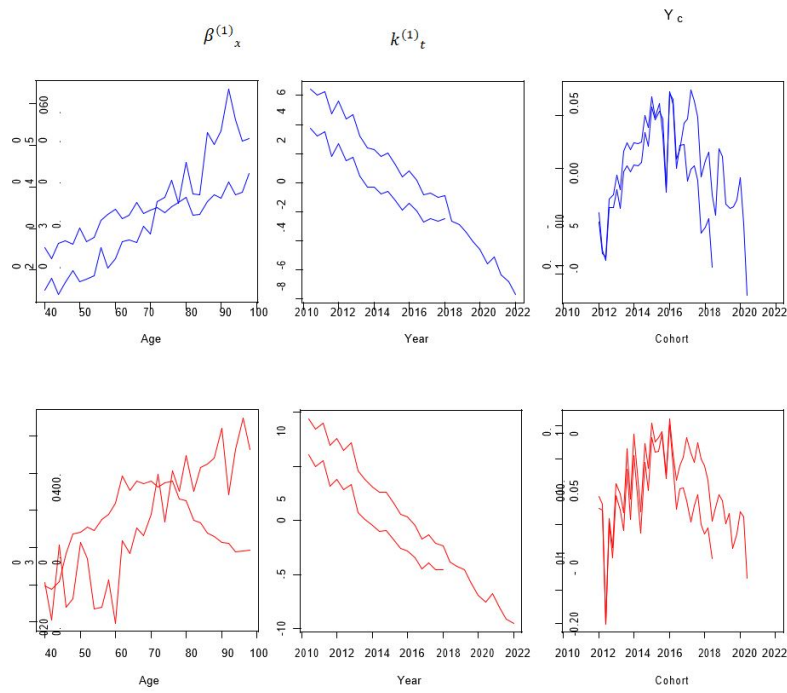
Figure 3.4.2: α_x, $\beta_x^{(1)}$ and $\kappa_t^{(1)}$ parameters estimates for males (blue top panels) & females (red bottom panels) for personal ages fitted from 2010 to 2022 for Model B

On the fitting of the data to the model, we start with model B, which is fitted as illustrated in Figure (3.4.2)
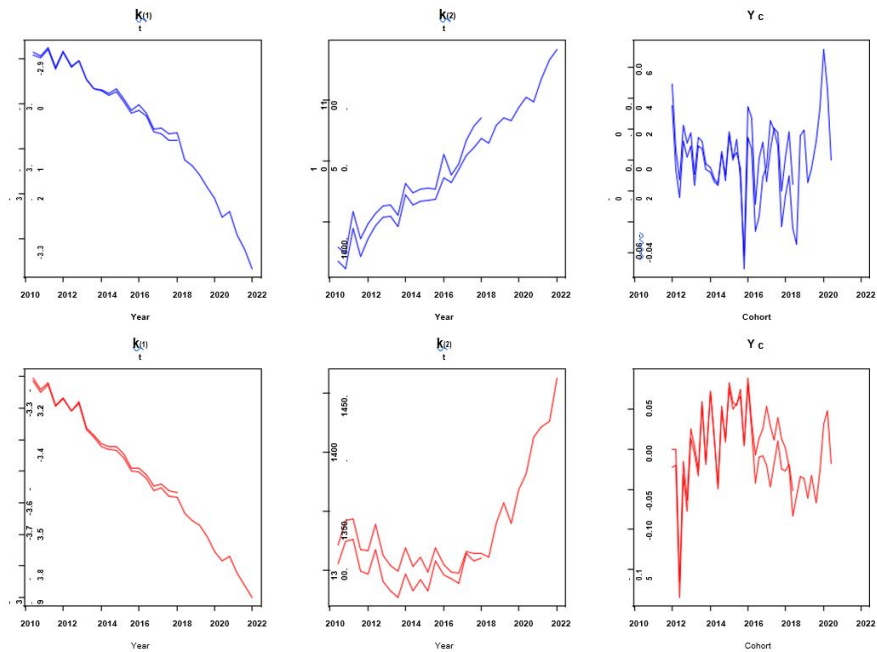
Figure 3.4.3: α_x, $\beta_x^{(1)}$ and $\kappa_t^{(1)}$ parameters estimates for males (blue top panels) & females (bottom panels) for personal ages fitted from 2010 to 2022 for Model C

22

3.4.1 Parameter Estimation and Robustness Property

The estimates of α_x in Figure 3.4.1 illustrates an approximately linear upward trend for SMR models for both genders. The same trend is almost similar for genders in models A, B, and C as shown in the Figures 3.4.2 to 3.4.3.

The $k_t^{(1)}$ estimates will always decrease in all mortality models, which indicates a general mortality index improvement for males and females over the specified time. For all the three models, which incorporate the cohort parameter, we cannot interpret that the model estimates cannot depend on the parameter estimations. It includes possible $k_t^{(2)}$ parameters interactions with respective corresponding effects of age. More precisely, cohort estimates of B (Figure 3.4.2), and C (Figure 3.4.3) are showing an upward trend in its increase.

As shown by Blake and Cairns (2021b), the robustness of the model is an important model's property that shows the sensitivity of estimates of parameter changes to the range of respective fitted data. It means that the parameter estimates need not change expressively whenever fitting data into a much shorter data range. Subsequently, probable inadequate robustness for a specific model does have a meaning of sensitivity to changes during the fitted data period while questioning the correctness of the use during projections or even many common applications, which independently rely on them.

Nevertheless, using a fitting range with fewer data may lead to a sudden increase in the number of $\beta_x^{(1)}$ for female estimates as illustrated in (bottom-left panel of Figure 3.4.4) while remaining unchanged even whenever the model is repeated for fitting bearing in mind fewer cohorts that should be excluded. On the other hand, models B (Figure 3.4.5) and C (Figure 3.4.6) appear to have the highest level of robustness for both males and females.
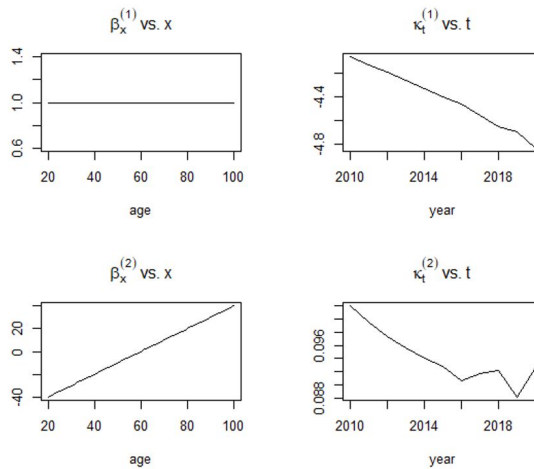
Figure 3.4.4: **A**: The Estimated parameters of $\alpha_x, k_t^{(1)}, \beta_x^{(1)}$ for Males fitted aged 20–100
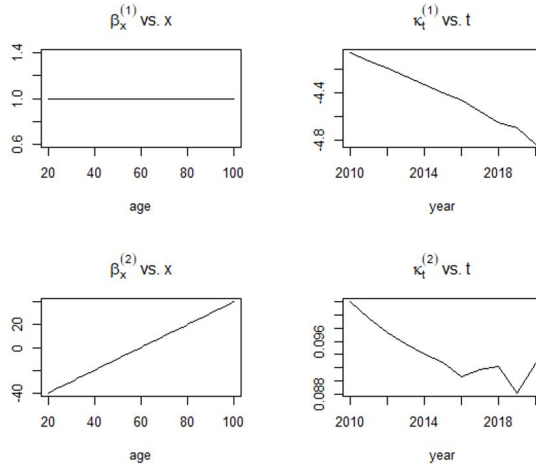
Figure 3.4.5: **B:** $k_t^{(1)}$, $\beta_x^{(1)}$, $k_t^{(2)}$ and γ_{w-x} estimated parameters for Males, aged 60–100
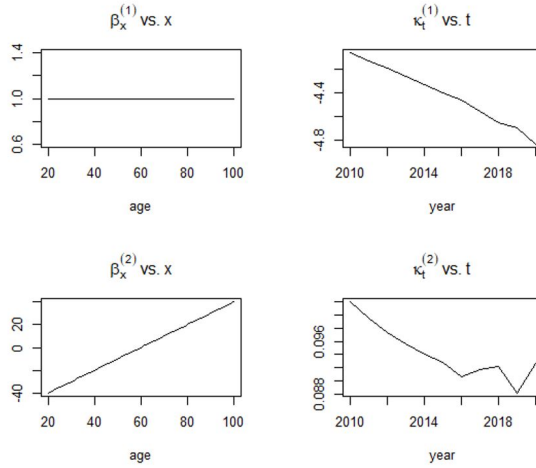
Figure 3.4.6: **C:** $k_t^{(1)}$, $\beta_x^{(1)}$ and γ_{w-x} estimated parameters for Males aged 60–100

3.4.2   Diagnostics of Goodness of Model's Fit

The residual differences between the fitted and observed data depend primarily on the chosen statistical distribution under the assumption and measuring the goodness of fit of the corresponding SMR model. Cupido, Fotheringham, and Jevtic (2021) in the study discussed that insufficient randomness, especially in the residuals patterns, can show the model's incapability to capture the precise age, time, and effects of cohorts.

Definition 3.4.1. Let $Z(\Theta, \hat{\Theta})$ under Binomial Distribution assumption that deaths be of residual deviance for every model defined as follows;

$$Z(\Theta, \hat{\Theta}) = \sum_{\forall x,t} Deviation(x,t) \tag{3.4.2}$$

where $Z(\Theta, \hat{\Theta}) = \sum_{\forall x,t} 2\varpi_{x,t} \left(\Theta \times ln(\frac{\Theta}{\hat{\Theta}}) + [\theta^0(x,t) - \Theta] \times ln(\frac{\theta^0(x,t) - \Theta}{\theta^0(x,t) - \hat{\Theta}}) \right), d(x,t) = \Theta, \widehat{d}(x,t) =$
$\hat{\Theta}$, and $\theta^0(x,t)$ is the initial exposure of the specified population. We obtain the standardized deviation defined as;

$$i_{x,t} = sign(\Theta, \hat{\Theta}) \times \left(\frac{Deviation(x,t)}{\hat{\phi}} \right)^{0.5} \tag{3.4.3}$$

With weights $\varpi_{x,t}$ in (3.4.3) are defined as;

$$\phi = \left(\frac{D(\Theta, \hat{\Theta})}{v} \right) \tag{3.4.4}$$

with v that expresses the model's degrees of freedom during modeling (the total number of the observations less number of the parameters a model has during estimation) and D is Deviation. The respective diagnostics of goodness of different SMR Models are fitted as;
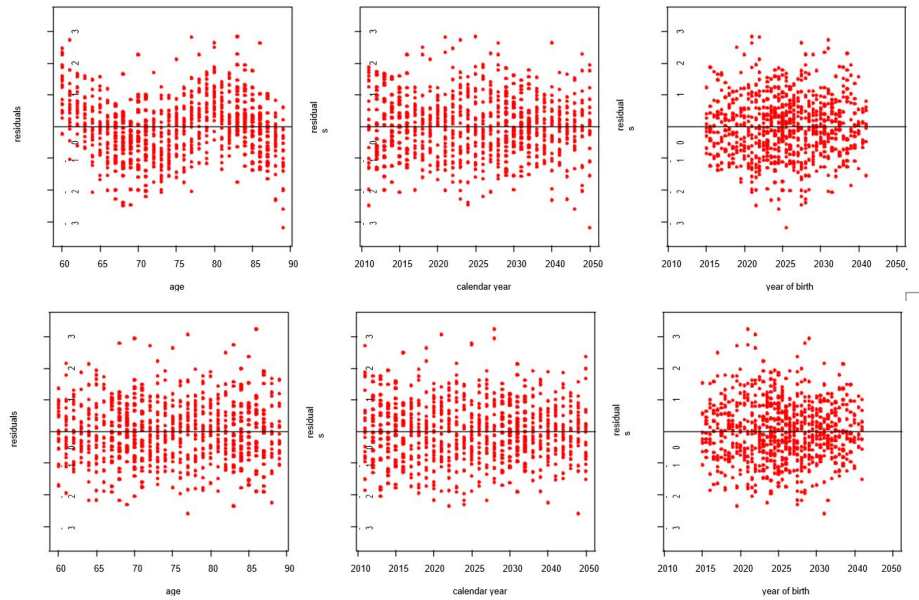
Figure 3.4.7: **A:** Residuals of Deviance for males (top consoles) & females (bottom consoles) for duration 2010-2020 from ages 60–100 for Kenya

Figure 3.4.7 show how residuals differences for Model A in the Kenyan population.
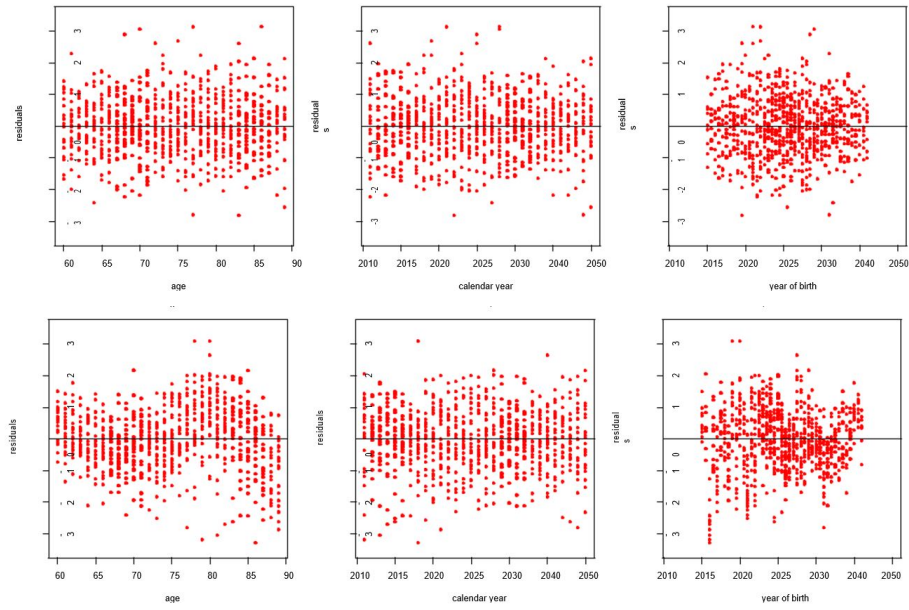
Figure 3.4.8: **B:** Residuals of Deviance for males (top consoles) & females (bottom consoles) for duration 2010-2020 from ages 60–100 for Kenya

Figure 3.4.8 show how residuals differences for Model B in the Kenyan population.
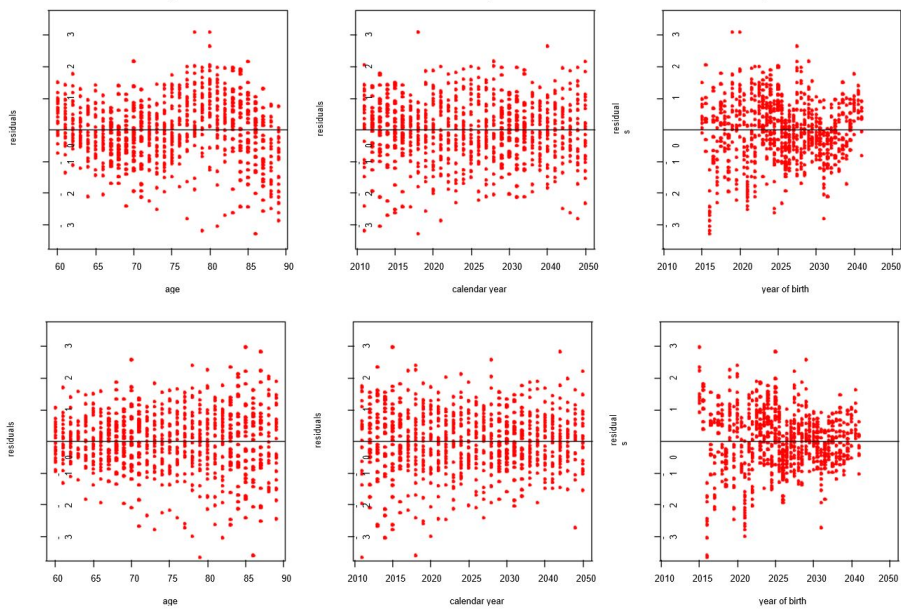
Figure 3.4.9: **C:** Residuals of Deviance for males (top consoles) & females (bottom consoles) for duration 2010-2020 from ages 60–100 for Kenya

Figure 3.4.9 show how residuals differences for Model C in the Kenyan population.

3.4.3 Information Criteria for Parameters

In general, with more parameters, it is possible to achieve a better fit. From the findings of Haberman and Renshaw (2011) that offered a substitute approach to addressing the assumption is to chastise the model parameters from the AIC method as in Akaike (1974) as well as BIC Brunton and Kutz (2022), which is the information criterion used for every model. Moreover, derivation of the Akaike criterion correction called the AIC(c) that works well when dealing with smaller samples was developed by Hurvich and Tsai (1989) and applied in statistical sample data by Burnham and Anderson (2004).

Definition 3.4.2. Let the Information Criteria denoted as AIC, *BIC*, and AIC(c) for the models; A, B, and C respectively be defined as;

$$AIC_{model=i}(c) = AIC_{model=i} + \left(\frac{2h_i(h_i+1)}{(n-h_i-1)}\right) \text{ with } AIC_{model=i} = 2h_i - 2ln\hat{L}_i$$

and

$$BIC_{model=i} = (ln\ n)h_i - 2ln\hat{L}_i$$

where the $\hat{L}_i$ is defined as the MLE and h_i are the number of parameters that has been determined by each of the 3 models while n is the sum number of observations during the parameter estimation and $i = A, B$ or C respectively.

When AIC, BIC, and AIC(c) respectively values are smaller, the model offers better fitting after estimations. For all the model's A, B, and C, Table 3.2 shows how these parameters fit in the respective SMR models for males, and Table 3.3 shows for females, respectively.

	Males			
Model	**Log Likelihood**	**AIC**	**BIC**	**AIC(c)**
A	−3522.925	6112.242(3)	6182.565(3)	6106.242(3)
B	−3245.225	6066.925(2)	6108.236(2)	6096.110(2)
C	−3672.345	6045.556(1)	6099.540(1)	6066.665(1)

Table 3.2: The Log Likelihood and BIC, AIC(c), and AIC values (order of ranking within brackets) of the SMR models for males

	Females			
Model	**Log Likelihood**	**AIC**	**BIC**	**AIC(c)**
A	−4140.832	7238.485(3)	7462.553(3)	7116.120(3)
B	−4142.235	7219.182(2)	7455.431(2)	7109.956(2)
C	−4099.485	7215.151(1)	7439.728(1)	7099.566(1)

Table 3.3: The Log Likelihood and BIC, AIC(c), and AIC values (order of ranking within brackets) of the SMR models for females.

Model C is on top using the BIC male results, followed by B, and A is third, whereas both AIC as well as AIC(c) male rankings concur. It is essential to note that BIC penalizes model parameters stronger than AIC(c) and AIC. Also, All the information criteria concur with A in terms of rank order for females. Unexpectedly, A models grasp the worst criteria ranking for both males and females, which indicates that cohort effect must account for male as well as female SMR modeling in Kenya.

3.4.4   Testing for Likelihood-Ratio

In Table 3.4 and 3.5, respectively, it is simple to observe that all models, *A*, *B*, and *C*, are different from others in unique ways. We use the LR test. With three tested models with their respective statistics illustrated for both males and females in Tables 3.4 and 3.5, respectively.

The LR statistic is given by $\Psi^{LR} = 2ln\left(\frac{\hat{L}_2}{\hat{L}_1}\right)$, where $\hat{L}_2$ is the MLE of the overall model and $\hat{L}_1$ of the nested model, whereas Ψ^{LR} is approximated as a χ^2 distribution, with $(n_2 - n_1)$ degrees of freedom, where n_2 are the degrees of the overall model and n_1 of the nested model. The null hypothesis is rejected for every pair of models in a level of significance α, as $\Psi^{LR} > \chi^2_{(n_2-n_1)}$, α that has a p-value of $1 - F\chi^2_{(n_2-n_1)}\left(\Psi^{LR}\right)$.

From the testing results, we confirm information criteria rankings, signifying that our models with less parameters fit on Kenyan data and are better than the more parsimonious SMR models.

H_0: General Model	H_1: Standard Model	Males LR of Test Statistic	Degrees of Freedom	p-Value
A	B	694.135	42	<0.0001
B	C	594.535	45	<0.0001
C	A	995.138	65	<0.0001

Table 3.4: LR test statistics for General models (H_0) within Specific models (H_1) for Males

H_0: General Model	H_1: Standard Model	Females LR of Test Statistic	Degrees of Freedom	p-Value
A	B	1259.455	42	<0.0001
B	C	1095.735	45	<0.0001
C	A	859.568	65	<0.0001

Table 3.5: LR test statistics for General models (H_0) within Specific models (H_1) for Females

3.5 Systematic Mortality Risk Projection

This subsection estimates the future mortality rates using models A, B, and C, for both males and females. Methods of projections are built on the extrapolation of the cohort with time (period) parameters for every model fitted on Kenyan mortality data. According to Currie (2016), selecting the best time series model can help in appropriate mortality modeling capable of reflecting both time(period) and cohort effects for a particular population with different characteristics. A comparative study was done by Hunt and Blake (2021) and Haberman and Renshaw (2011) modeled period indices through a multivariate random walk having both drift as well as cohort indices using univariate ARIMA models.

In this specific case, we have chosen the correct univariate ARIMA model for every cohort index and the period over a wide range of our models, Choi (2001), Dickey and Fuller (1979) and unit root tests, as well as the information criteria values. To be more specific, the selection was based on the general performance of time series against the penalized scores of AIC, BIC, and AICc. If one prefer simpler time series models on the parsimony illustrates the discordance issues in between the criteria values. Thus, κ_t's in models A, B, and C, are presumed to be independent of the equivalent W_{t-c} for every of the SMR model, following correspondingly univariate ARIMA(p,d,q) processes that takes on the forms;

$$\left(1 - \phi_1 Z - \phi_2 Z^2 - ...\phi_p Z^p\right)(1-Z)^d k_t = \delta + \left(1 + \vartheta_1 Z + \vartheta_2 Z^2 + ... + \vartheta_q Z^q\right) e(t) \tag{3.5.1}$$

$$\left(1 - \phi_1' Z - \phi_2' Z^2 - ...\phi_p' Z^p\right)(1-Z)^d W_{t-c} = \delta' + \left(1 + \vartheta_1' Z + \vartheta_2' Z^2 + ... + \vartheta_q' Z^q\right) e'(t) \tag{3.5.2}$$

where Z^d is a time lag operator (sometimes called back-shift operator), which shifts data d periods back, δ' and δ are known as constant drift parameters, $\phi_1', \phi_2', ..., \phi_p'$ and $\phi_1, \phi_2, ..., \phi_p$ are known as the auto-regressive coefficients having $\phi_p \neq 0, \phi_p' \neq 0$, where as $\vartheta_1, \vartheta_2, ..., \vartheta_q$ and $\vartheta_1', \vartheta_2', ..., \vartheta_q'$ are the MA parameters with $\vartheta_p \neq 0, \vartheta_p' \neq 0$ and the values of $e(t)$ and $e'(t)$ are the White Noises.

Tables 3.6 and 3.7 present the chosen ARIMA models for both time (period) and cohort indices for both genders (males and females), respectively. For all models of mortality, time indices are presumed to be modeled as identically and independently distributed. In addition, remember that A and C do not integrate a cohort index as in the classical models. The two time series equations (3.5.1) and (3.5.2) were simulated after producing 500 trajectories for the future values of the specific period $\hat{k}_{t_n+s}$ and the cohort $\hat{W}_{t_n+s-x}$ indices, where $s = 1, 2, 3, \cdots, 20$ are the predicting horizon years.

Definition 3.5.1. Let $logit\hat{q}_{x,t_n+s}$ be the future simulated mortality values extracted using the following information be;

$$logit \hat{q}_{x,t_n+s} = \alpha_x + \sum_{i=1}^{N} \beta_x^{(i)} \hat{k}_{t_n+s}^{(i)} + W_{n+s-x}$$

which can also be defined as follows;

$$\hat{q}_{x,t_n+s} = \left(\frac{e^{\alpha_x + \sum_{i=1}^{N} \beta_x^{(i)} \hat{k}_{t_n+s}^{(i)} + W_{n+s-x}}}{1 + e^{\alpha_x + \sum_{i=1}^{N} \beta_x^{(i)} \hat{k}_{t_n+s}^{(i)} + W_{n+s-x}}} \right) \tag{3.5.3}$$

where $t_n = 2020$ is the previous year of the fitting time and $logit \hat{q}_{x,t_n+s}$ is the *logit* transformation of future death probabilities for every exact age x for A, B, and C models.

We extracted both short-term forecast errors for males and females during 2010–2020 for A, B, and C models. For comparison and extrapolation, the fitted jump-off rates have been used before using actual rates for the year 2010, shown in Table 3.9. The results are then applied from the Kenyan Mortality data. Measures show that models A, B, and C, resulted in better forecasts for both males and females (ranking order in brackets), either through actual or fitted jump-off rates.

Particularly, when fitted values are used, models A and C differentiate for both males and females, whereas for actual rates, C is dominant for males while B outperforms others for females. The error measures produced the higher values of errors for A, and C, which indicates cohort effects in both gender mortality indices, which cannot be illustrated by model B and probable over-fitting behavior of C.

Model	Males $k_t^{(1)}$	$k_t^{(2)}$
A	ARIMA(0,0,1)	-
B	ARIMA(1,0,1) with drift	ARIMA(1,1,1) with drift
C	ARIMA(1,2,1)	-

Table 3.6: ARIMA(p,d,q) models for the time index $k_t^{(i)}$, $i = 1,2,3$ of males in SMR models.

Model	Females $k_t^{(1)}$	$k_t^{(2)}$
A	ARIMA(0,2,2)	-
B	ARIMA(0,0,1) with drift	ARIMA(2,1,0) with drift
C	ARIMA(1,2,1)	-

Table 3.7: ARIMA(p,d,q) models for the time index $k_t^{(i)}$, $i = 1,2,3$ of females in SMR models

Model	W_{t-c} for Males	W_{t-c} for Females
A	ARIMA(1,1,0)	ARIMA(1,1,1)
B	ARIMA(1,1,1)	ARIMA(2,1,1)
C	ARIMA(1,1,0)	ARIMA(1,1,0)

Table 3.8: ARIMA(p,d,q) models for the cohort index W_{t-c} of female and male SMR models

It is vital to calculate the predictive power of these SMR models to measure the errors in between the recorded and predicted values for a similar period. For the initial three out-of-sample for projection years ($t_n = 2020, s = 1,2,3,...$), where we applied the Kenyan mortality data, forecast precision of models, *A*, *B*, and *C* has been determined through getting the average of the MAE as well as the MAPE values over the four years for ages between 60 to 100.

Definition 3.5.2. Let error measures be calculated as follows;

$$MAE_{mean} = \frac{1}{4(100-60+1)} \sum_{s=1}^{4} \sum_{x=60}^{100} (\hat{q}_{x,2020+s} - q_{x,2020+s}) * 100 \qquad (3.5.4)$$

while the

$$MAPE_{mean} = \frac{1}{4(100-60+1)} \sum_{s=1}^{4} \sum_{x=60}^{100} \left(\frac{\hat{q}_{x,2020+s} - q_{x,2020+s}}{q_{x,2020+s}} \right) * 100 \qquad (3.5.5)$$

From the above two equations, namely (3.5.4) and (3.5.5), we can determine the expected values of MAE and MAPE that measure the forecasting period 2020-2050 using either actual or fitted jump-off rates for both genders (males and females).

Males				Females			
Male-Fitted	**Jump**	**Off**	**Rates**	**Fitted**	**Jump**	**Off**	**Rates**
Error	**A**	**B**	**C**	**Error**	**A**	**B**	**C**
MAE_{mean}	0.332(3)	0.262(1)	0.281(2)	MAE_{mean}	0.239(2)	0.255(3)	0.229(1)
$MAPE_{mean}$	11.832(3)	10.450(1)	10.842(2)	$MAPE_{mean}$	8.328(3)	7.035(2)	6.834(1)
Actual	**Jump**	**Off**	**Rates**	**Actual**	**Jump**	**Off**	**Rates**
Error	**A**	**B**	**C**	**Error**	**A**	**B**	**C**
MAE_{mean}	0.265(3)	0.248(1)	0.256(2)	MAE_{mean}	0.239(3)	0.226(2)	0.217(1)
$MAPE_{mean}$	10.332(3)	9.232(1)	9.838(2)	$MAPE_{mean}$	9.950(3)	8.922(1)	9.436(2)

Table 3.9: Expected values (grading order in brackets) of MAPE and MAE of the predicting period 2010–2020 using fitted jump-off rates for Kenyans

The long-term systematic mortality risk projections for a horizon ahead of 30 years were determined using actual jump-off rates for the ten mortality risk models, which incorporated 500 simulation trajectories of the chosen period and cohort indices.

Figures (3.5.1-3.5.2) shows that A, B, and C forecasts appear to be improbable for both males and females since fans at age 90 are markedly narrower than at age 60. Additionally, B fans at age 80 show a weak but not substantial increase for both males and females, whereas fans at age 90 show some declining fluctuations. Alternatively, parsimonious model B performs well in general for both males and females. In conclusion, female fans of model A are narrower age 80 and 90 than at 60 and showing an unrealistic increase at older ages. Predictions are linked to the predictable cohort effect thus exhibiting a strong, vertical, and linear trend in between cohort years from 2010 - 2020.

3.5.1 Assessment of Parameter Risks

For those countries with limited data experience like Kenya, however, bootstrapping techniques can be used when addressing the effect of risk parameters. Thus, we exploited the benefits of a residual bootstrapping method when assessing the parameter uncertainty in the mortality projections for the three models.

Figures 3.5.1-3.5.2 illustrates for both males and females the 95% prediction intervals for the death probabilities at ages $x = 60$, $x = 80$, and $x = 100$ for models A, B, and C fitted to Kenyan data for ages 60–100 of the time 2010–2050.

The past rates are denoted using thick dots while solid lines indicate the corresponding fitted rates, while dot-dashed lines showing the 95% confidence intervals, including the unpredictability of the parameter. For the projection period of 2010–2050, dashed lines in the diagram represent the central forecast values, while dot lines show the 95% prediction intervals, which excludes parameter uncertainty. In addition, the dot-dashed lines showing the 95% prediction intervals levels for parameter unpredictability.

Figure 3.5.1 is also showing evidence of parameter unpredictability in the time of projection for males (left panels) of models B (age 90) and C (ages 70 and 90). In addition, parameter variability is seen for females of the same models (right panels). An improbable upward trend for C at age 80 and 90 indicates the model's unsuitability when forecasting female mortality at higher ages.
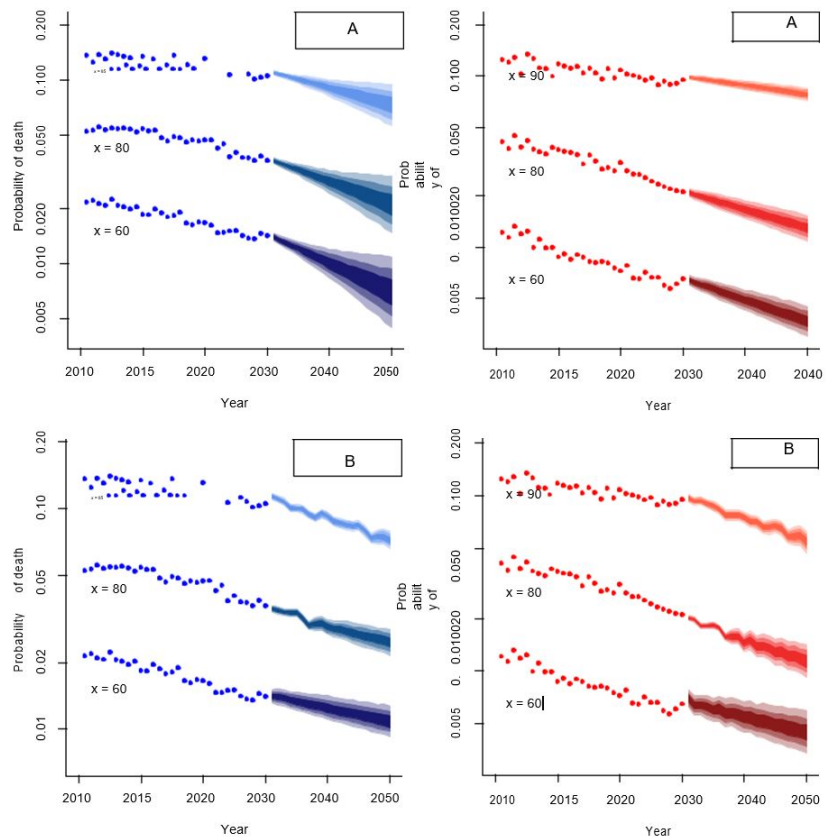
Figure 3.5.1: Long-term SMR prediction of A, and B models fitted from 2010 to 2020 and projections from 2020 to 2050 for ages 60 to 100 for both males & females for confidence levels of 50%, 80% and 95% intervals of prediction respectively
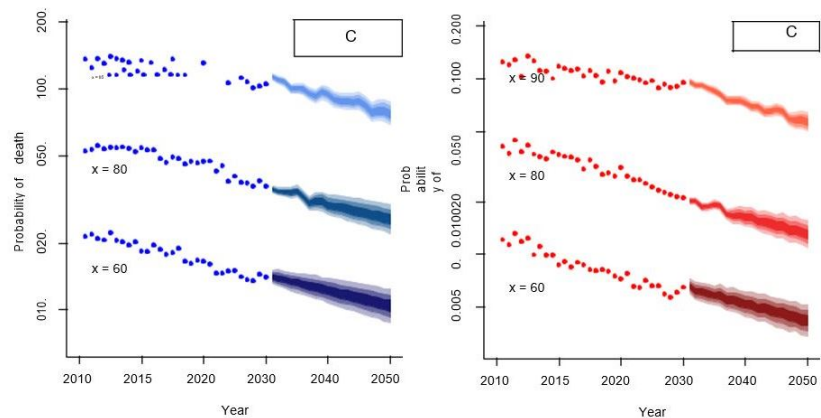
Figure 3.5.2: Long-term SMR prediction of C models fitted from 2010 to 2020 and projections from 2020 to 2050 for ages 60 to 100 for both males & females for confidence levels of 50%, 80% and 95% intervals of prediction respectively

3.6   Analysis and Results

This section summarizes the fitting and prediction results of this analysis and findings to compare the matching results from the classical mortality studies. This study proves that all the SMR models effectively captured the time effects for both males and females. It is vital to notice that the most parsimonious model is B. It did not capture well the cohort effects as demonstrated in the right panels of males and females scatter plots in terms of their residual deviance in Figures 3.4.8. Moreover, model A seems insufficient when capturing the effects of age, particularly for females (left consoles as shown in Figures (3.4.7).

Moreover, AIC and AIC(c) scores concur since models A and B do outperform in ordered ranking for males, whereas BIC rankings C is on top, B follows, and A is third. All measured values show that C comes out first, B second, and A follows for females. Models A and B have the worst criteria ranking for both genders, lacking a cohort term to consider in Kenyan male and female SMR modeling.

SMR projections got from the three models as illustrated for both genders in Figures (3.5.1) and (3.5.2). Through plotting, results depict that long-term forecasts from models A, B, and C seem to be undependable for both males and females since figures at age 90 are conspicuously narrower than at age 60. In addition, model A for females shows an improbable increase in SMR at ages 70 and 90. Nevertheless, forecast accuracy measures of Table 3.6 and Table 3.7 suggest that models B and C produce better short-term forecasts for both males and females.

Parameter inconsistency is also observed in model B for females. Similarly, the improbable upward trend for A at age 70 and 90 raises questions about the suitability of this model when forecasting the mortality of females in Kenya.

A comparative analysis is done from the three stochastic SMR models of a standard ATC structure for males and females in Kenya. The fitting behavior of every model was calculated using BIC, AIC, and AIC(c), respectively, and the likelihood ratio test and the respective forecasting results were presented. The models, B, and C for males and B and A for females, were differentiated for their fitting performances during mortality modeling. For the Kenyan case, a cohort effect was determined from data taken when choosing the most appropriate model for modeling SMR.
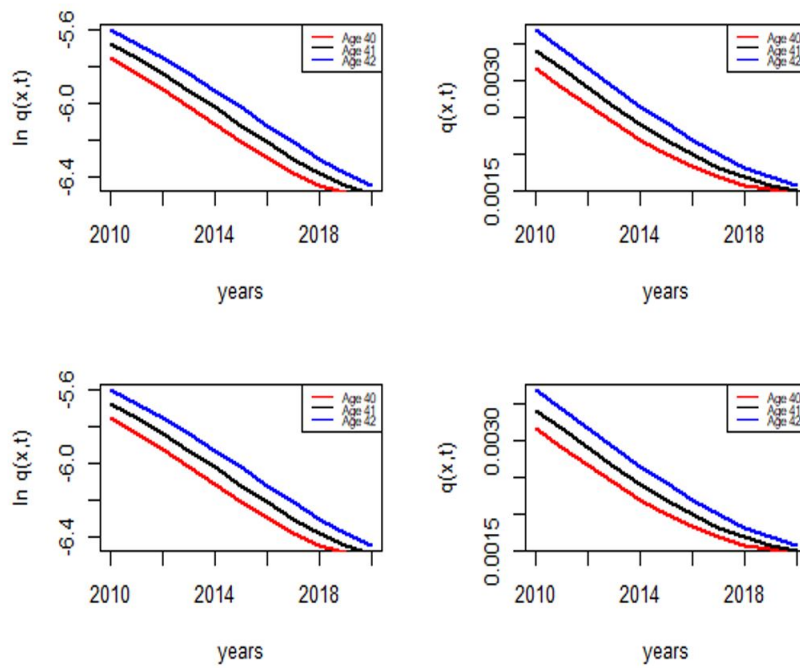
Figure 3.6.1: **A**(up) and **B**(down) for females(right) & males(left)
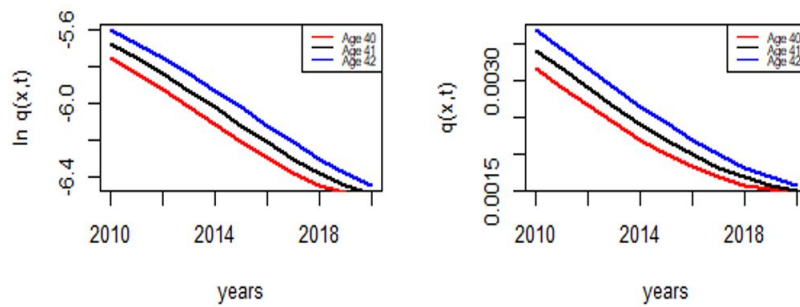
Figure 3.6.2: **C** model for females(right) & males(left)

Finally, the plots from Figures 3.6.1 and Figure 3.6.2 demonstrates the overall evolution of death numbers for every of the SMR model under the ATC structure.

Chapter 4

Bühlmann Credibility Approach Incorporation into Systematic Mortality Risk Modeling

In this chapter, we integrated the Bühlmann credibility approach into SMR modeling. We introduced the Bühlmann Credibility model description and applications in SMR modeling. We do model modification on the classical SMR models under 3-factor modelling described in Chapter 3 by proposing to model the randomness of the Stochastic Mortality models as NIG distribution (heavy-tailed distribution) since the randomness of the classical models does not exhibit the Gaussian assumptions in the Kenyan setup due to data paucity.

We measured MAPE, RMSE, and reduction ratios of the three mortality models under Bühlmann's credibility approach to Kenyan mortality data. We explain the empirical illustration using the Kenyan data for males and females, then presented with forecasting performances of Bühlmann credibility. The models are calculated with the measures of MAPE and RMSE before fitting.

We noted the new predicting capabilities of the SMR models under the Bühlmann credibility approach from the risk measures. We did a comparison between these three models by conducting an actuarial valuation.

4.1   Bühlmann Credibility Model Description

Let consider a risk i that generates random losses for the historical data of m recent deaths in our case are available and indexed by j . The total deaths experienced for i the risk is

36

determined based on the expected deaths that exist during a specified period in mortality studies see Klugman, Panjer, and Willmot (2012). We use a linear estimator that diminishes the mean square error in the study.

Definition 4.1.1. Let Y_{ij} be the ith risk for the *jth* death, which are *iid* with $\bar{Y} = \sum_{i=1}^{n} \sum_{j=1}^{m} Y_{ij}$. Let Θ_i denote the parameter associated with the distribution of the SMR. This means the hypothetical mean given by $m(\vartheta) = \mathbb{E}\left[Y_{ij}|\Theta_i = \vartheta\right]$ and process variance is $s^2(\vartheta) = Var\left[Y_{ij}|\Theta_i = \vartheta\right]$. In addition, let $\Pi = \mathbb{E}[(m(\vartheta)|Y_{i1}, Y_{i2}, ...Y_{im})]$ be the parameter for the SMR, with $\sigma^2 = \left[s^2(\vartheta)\right]$ and $a^2 = Var\left[m(\vartheta)\right]$ being the two functions of the random parameter ϑ.

Lemma 4.1.1. *Let the BC Model predictor be a solution of the estimator problem as in Klugman, Panjer, and Willmot (2012) then*

$$\underset{v_{i0}, v_{i1}, ..., v_{im}}{argmin} \ \mathbb{E}\left[\left(v_0 + \sum_{i=1}^{n} \sum_{j=1}^{m} v_{ij}Y_{ij} - \Pi\right)^2\right]$$

where $v_0 + \sum_{i=1}^{n} \sum_{j=1}^{m} v_{ij}Y_{ij}$ is the estimator of expected SMR, Π, and arg min component represents the expression of parameter values that diminishes the function for all $i \neq j$.

Then the BC Model solution is given by;

$$Z\bar{Y} + (1 - Z)\mu \tag{4.1.1}$$

where Z is such that $Z = \left(\frac{n}{n+K}\right)$ and $K = \frac{\sigma^2}{a_m^2}$. Therefore, making the value of K small and making Z closer to the value 1. Another attractive feature of the BC formula is that the more the amount of experience data accumulated (as $n \to \infty$), the more the credibility factor of Z approaches one. It is vital to note that equation (4.1.1) is essential in application of the BCA to SMR modeling as it incorporates all used linear estimators of the number of deaths experienced in a specific population.

Lemma 4.1.2. *Let BCA problem of Mortality Quadratic Loss be stated in a counteractive way to satisfy equation (4.1.1) for all $i \neq j$, then;*

$$f = \mathbb{E}\left[\left(v_0 + \sum_{i=1}^{n} \sum_{j=1}^{m} v_{ij}Y_{ij} - m(\vartheta)\right)^2\right] \to min \ C = Z\bar{Y} + (1 - Z)\mu \ where \ C$$

is the Credibility Estimate for Systematic Mortality Risk.

Proof. We verify by assuming that the estimator for the individual mean death will be $m(\vartheta)$ and simple BC is $\mathbb{E}\left[\left(v_0 + \sum_{i=1}^{n} \sum_{j=1}^{m} v_{ij}Y_{ij} - m(\vartheta)\right)^2\right]$ $\qquad \qquad \square$

$$= \mathbb{E}\left[\left(v_a + \sum_{i=1}^{n} \sum_{j=1}^{m} vaY_{ij} - \Pi\right)^2\right] + \mathbb{E}\left[(m(\vartheta) - \Pi)\right] + 2\mathbb{E}\left[\left(v_0 + \sum_{i=1}^{n} \sum_{j=1}^{m} v_{ij}Y_{ij} - \Pi\right)(m(\vartheta) - \Pi)\right]$$

$$= \mathbb{E}\left[\left(v_0 + \sum_{i=1}^{n}\sum_{j=1}^{m} v_{ij}Y_{ij} - \Pi\right)^2\right] + \mathbb{E}\left[(m(\vartheta) - \Pi)^2\right] \qquad (4.1.2)$$

Expanding the equation (4.1.2) and collecting the terms it becomes;

$$= \mathbb{E}\left[\left(v_0 + \sum_{i=1}^{n}\sum_{j=1}^{m} v_{ij}Y_{ij} - \Pi\right)(m(\vartheta) - \Pi)\right] \qquad (4.1.3)$$

$$= \mathbb{E}_\Theta\left[\mathbb{E}_Y\left[\left(v_0 + \sum_{i=1}^{n}\sum_{j=1}^{m} v_{ij}Y_{ij} - \Pi\right)(m(\vartheta) - \Pi)\Bigg| Y_1, Y_2 \ldots, Y_m\right]\right]$$

$$= \mathbb{E}_\Theta\left[\left(v_0 + v_{ij}Y_{ij} - \Pi\right)\left[\mathbb{E}_Y\left[(m(\vartheta) - \Pi)|Y_1, Y_2 \ldots, Y_m\right]\right]\right]$$

$$= 0$$

We apply the aggregate expectation and the fact, that $\Pi = \mathbb{E}[m(\vartheta)|Y_1, Y_2 \ldots, Y_n]$.

From the equation (4.1.3), we decompose its minimized function as an aggregate of the two expressions where its second expression doesn't depend on the parameters that have been used in the minimization. Thus, minimizing the function is done in a similar manner to minimizing the first part of the aggregate sum.

By minimization of its function critical points, we obtain;

$$\frac{1}{2}\frac{\partial f}{\partial v_{01}} = \mathbb{E}\left[v_0 + \sum_{i=1}^{n}\sum_{j=1}^{m} v_{ij}Y_{ij} - m(\vartheta)\right]$$

$$= v_0 + \sum_{i=1}^{n}\sum_{j=1}^{m} v_{ij}\mathbb{E}(Y_{ij}) - \mathbb{E}(m(\vartheta))$$

$$= v_0 + \left(\sum_{i=1}^{n}\sum_{j=1}^{m} v_{ij} - 1\right)\mu \text{ for all values of } k = 0.$$

For the values of $k \neq 0$, we have:

$$\frac{1}{2}\frac{\partial f}{\partial v_{ik}} = \mathbb{E}\left[Y_k\left(v_0 + \sum_{i=1}^{n}\sum_{j=1}^{m} v_{ij}Y_{ij} - m(\vartheta)\right)\right]$$

$$= \mathbb{E}[Y_k]v_0 + \sum_{i=1}^{n}\sum_{j=1, j\neq k}^{m} v_{ij}\mathbb{E}[Y_{ik}Y_{ij}] + v_k\mathbb{E}[Y_{ik}^2] - \mathbb{E}[Y_{ik}m(\vartheta)] = 0$$

We then simplify its derivative, by noting that it has:

38

$$\mathbb{E}[Y_{ij}Y_{ik}] = \mathbb{E}\left[\mathbb{E}[Y_{ij}Y_{ik}|\vartheta]\right]$$

$$\mathbb{E}[Y_{ik}^2] = \mathbb{E}\left[\mathbb{E}[Y_{ik}^2|\vartheta]\right] = \mathbb{E}[s^2(\vartheta) + (m(\vartheta))^2] = \sigma^2 + p^2 + \mu^2$$

$$\mathbb{E}[Y_{ik}m(\vartheta)] = \mathbb{E}[\mathbb{E}[Y_{ik}m(\vartheta)|\Theta_i] = \mathbb{E}[(m(\vartheta))^2] = p^2 + \mu^2 \qquad (4.1.4)$$

Taking above equation (4.1.4), and inserting into derivative, it gives:

$$\frac{1}{2}\frac{\partial f}{\partial v_k} = 0$$

$$\left(1 - \sum_{i=1}^{n}\sum_{j=1}^{m}v_{ij}\right)\mu^2 + \sum_{i=1}^{n}\sum_{j=1,j\neq k}^{m}v_{ij}\left(p^2 + \mu^2\right) + v_k\left(\sigma^2 + p^2 + \mu^2\right) - \left(p^2 + \mu^2\right) = 0$$

which ultimately becomes

$$v_k\sigma^2 - \left(1 - \sum_{i=1}^{n}\sum_{j=1}^{m}v_{ij}\right)p^2 = 0$$

$$\sigma^2 v_k = p^2\left(1 - \sum_{i=1}^{n}\sum_{j=1}^{m}v_{ij}\right)$$

The right hand side of the equation doesn't depend on the value of k. Therefore, all v_k are constant. It follows;

$$v_1 = v_2 = v_3 \cdots = v_m = \frac{p^2}{\sigma^2 + mp^2}$$

From the solution for v_{i0}, we have;

$$v_0 = (1 - mv_k)\mu = \left(1 - \frac{mp^2}{\sigma^2 + mp^2}\right)\mu$$

Finally, the best estimator is

$$\left(v_0 + \sum_{i=1}^{n}\sum_{j=1}^{m}v_{ij}Y_{ij}\right) = \left(\frac{mp^2}{\sigma^2 + mp^2}\right)\overline{Y} + \left(1 - \frac{mp^2}{\sigma^2 + mp^2}\right)\mu$$

$$C = Z\overline{Y} + (1 - Z)\mu$$

where C is the BC Estimate.   Hence, the proof.

4.2 Heavy Tailed Distribution

4.2.1 Introduction

We had reviewed the SMR models by using **Definition 3.1.1.**, **Definition 3.1.2.** and **Definition 3.1.3.** as summarized in **Table 3.1**. The randomness of these classical models follows a Normal distribution, which does not fit well, as illustrated in **Chapter 3**. We model the disturbances of these classical SMR models as a heavy-tailed distribution. We begin by introducing the concept of heavy-tailed distribution.

Definition 4.2.1. Let $f(Y)$ be the PDF of a heavy tailed distribution function of a random variable Y if its MGF is infinite for all values of t, y greater than 0, then

$\int\limits_{-\infty}^{\infty} e^{ty} f(Y) dy = \infty$

From definition 4.2.1, it implies that;

$$\lim_{y \longrightarrow \infty} e^{ty} P[Y > y] = \infty \ \forall \, t > 0$$

For the Complementary CDF, $S(y) = Pr[Y > y] = 0 \le S(y) \le 1$ and $\lim\limits_{y \longrightarrow \infty} e^{ty} S(y) = 0 \le S(y) \le 1$ for all values of $t > 0$.

where $S(y)$ is the survival function of Y, which is obtained by $S(y) = 1 - F(y)$ and $F(y)$ is the CDF of Y.

Definition 4.2.2. Let a heavy tailed statistical distribution be in the form of;

$$\int\limits_{\mathbb{R}} e^{\lambda y} F(dy) = \infty \ \forall \text{ values of } \lambda > 0$$

where the function $g(y) \ge 0$ is to be a heavy tailed iff;

$$\lim_{y \longrightarrow \infty} g(y) e^{\lambda y} = 0 \ \forall \text{ values of } \lambda > 0.$$

Theorem 4.2.1. *Let G be a function of a heavy-tailed statistical distribution with a fixed time, $T > 0 \ \forall \ G(y, y + T]$, then the corresponding force of systematic mortality (hazard function) satisfies the function $\lim\limits_{y \longrightarrow \infty} \inf \left(\frac{R(y)}{y} \right) = 0.$*

Proof. Proving by contradiction method. Suppose we assume that the function $G(y, y + T]$ is not heavy tailed. Then, we can define;

$d := \sup G(y, y + T] e^{\lambda' y} < \infty$ for some values of $\lambda' > 0$. Therefore, $\forall \lambda < \lambda'$, we get;

$$\int\limits_{0}^{\infty} e^{\lambda y} G(dy) \le \sum_{m=0}^{\infty} e^{\lambda (m+1)T} G(mT, mT + T]$$

$$\leq d \sum_{m=0}^{\infty} e^{\lambda(m+1)T} e^{-\lambda'mT} = d\, exp(\lambda T) \sum_{m=0}^{\infty} e^{mT(\lambda-\lambda')} < \infty$$

From the above integral defined where it is determinate for all values of $\lambda \in (0, \lambda')$ that has an implication on the distribution of G and cannot be a heavy tailed distribution. Thus the needed implication is shown as;.

We define that the implication follows from the specific inequality $G(y) \geq G(y, y+T]$.

Assume that on the contrary of the proposal, '$lim\,inf$' as defined as a strictly positive number (> 0).

Then, we say there exist $y_0 > 0$ and $\varepsilon > 0$ such that $R(y) \geq \varepsilon y$ for all $y \geq y_0$ which implies that $G(y) \leq e^{-\varepsilon y}$ thus a contradiction.

Assume that, on the contrary, G is *NOT* a heavy tailed distribution, then we deduce (for instance by the exponential Chebyshev inequality) that, for some values of $\lambda > 0$ and $d > 0$, we have $G(y) \leq de^{-\lambda y}$ for all y.

This implies that $\lim_{y \longrightarrow \infty} inf\left(\frac{R(y)}{y}\right) = 0$ thus contradiction.

This leads to the conclusion that the distribution is heavy tailed.

Hence, the proof. □

Lemma 4.2.1. *Let $f_Y(y)$ be a NIG distribution with the Tail behavior that is assumed to be a Bessel function that asymptotically behaves as;*

$$K_{\frac{1}{2}}(y) \sim \sqrt{\frac{\pi}{2y}} e^{-y}; |y| \longrightarrow \infty$$

Proof. Let us note that the tail of Normal Inverse Gaussian distribution decays as

$$f_Y(y) \sim |y|^{\frac{-3}{2}} e^{\beta y - \alpha|y|} \tag{4.2.1}$$

From the above equation (4.2.1), if it is invalid, then the value of $\alpha - |\beta| < 1$. Through this, a special case of the *NIG* tail decays as

$$f_Y(y) \sim |y|^{-2}$$

and this is a tail behavior of the Cauchy distribution and hence, a heavy tailed distribution.

Hence, the proof. □

4.2.2 Normal Inverse Gaussian (NIG) Distribution

The NIG distribution is a continuous distribution function defined as a normal variance-mean mixture with a mixing density as the IG distribution. It is Generalised Hyperbolic

Distribution a special case with four parameters, namely α, β, θ, and σ defined as tail heaviness, asymmetry parameter, location, and scale parameter, respectively, all of which are real numbers ($\alpha > 0$ and $\sigma > 0$).

Definition 4.2.3. The random variable Y is said to have Normal Inverse Gaussian distributed NIG $(\alpha, \beta, \theta, \sigma)$ if its PDF is given by;

$$f_y(NIG) = \frac{\alpha}{\pi} exp\left\{\sigma\sqrt{\alpha^2 - \beta^2} - \beta\theta\right\} \phi(y)K_1\left(\sigma\alpha\phi(y)^{\frac{1}{2}}\right)exp(\beta y) \qquad (4.2.2)$$

where $\phi(y) = \left[1 + \left(\frac{(y-\theta)}{\sigma}\right)^2\right]$, $K_r(y)$ is the modified Bessel function the $3rd$ kind of order r calculated at y. The conditions for the four parameters are $0 \leq |\beta| \leq \alpha, \theta \in \mathbb{R}$, and $0 < \sigma$. Different parameters on the distribution play unique roles that can be attributed to two groups. The first group denotes α and β that mainly affects the shape of the distribution. The second group of parameters belongs to $\mu = \theta$ and $\delta = \sigma$ that denotes distribution scale. The parameter α, which can take non-negative values, denotes density function flatness. It means the greater value of the α, the greater the probability mass concentration around θ. In addition, the probability density function will reach a much higher maximum value; look at Figure 4.2.1(a).

On the other side, parameter β is determining the kind of skewness the distribution posses. Whenever the value $\beta = 0$, it implies that the symmetric distribution is around the mean. However, a negative value of β means a heavier left tail, and a positive value means a heavier right tail (see Figure 4.2.1(b). Moreover, the 3rd parameter $\delta = \sigma$ denoting the distribution scale. It means that small values of σ narrow the distribution down while larger values of σ make it wider. The last parameter θ is responsible for the density function shift, which is the location parameter.

(a) varying α and $\beta = 2, \mu = -3, \delta = 1$ (b) varying β and $\alpha = 1, \mu = 0, \delta = 1$
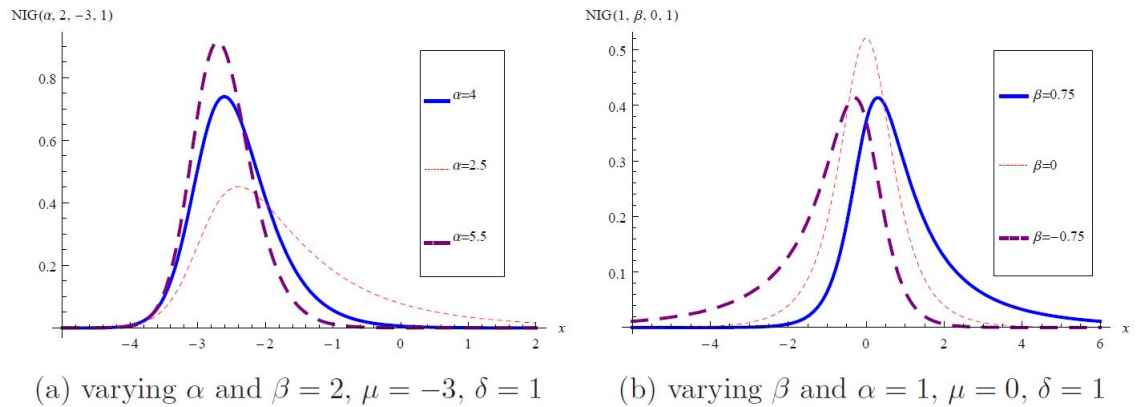
Figure 4.2.1: **NIG Probability Density plots**

The NIG $(\alpha, \beta, \theta, \sigma)$ distribution is always defined as a Gaussian Variance-Mean mixture, which means that the distribution can be given out as the marginal distribution of variable Y in the pair (Y, Z), with the conditional probability $Y|Z$ defined by

$$Y/Z = Z \sim N(\theta + \beta z) \tag{4.2.3}$$

and the variable Z follows IG distribution with the parameters σ and $\sqrt{\alpha^2 - \beta^2}$ for $0 \leq |\beta| \leq \alpha$.

Definition 4.2.4. Let denote $\tau = \sqrt{\alpha^2 - \beta^2}$, then the mean, variance, kurtosis and skewness of Y be given by;

$$Y \sim \begin{cases} m = & \theta + \frac{\beta}{\tau}\sigma \\ v = & \sigma\frac{\alpha^2}{\tau^3} \\ s = & 3\frac{\beta}{\alpha(\sigma\tau)^{0.5}} \\ k = & 3\frac{(1+4(\frac{\beta^2}{\alpha^2}))}{\sigma\tau} \end{cases}$$

where mean is $\mathbb{E}[y] = m$, variance$= Var(y) = v$, kurtosis$= s$, and skewness$= k$.

Despite the NIG having a fairly complicated looking probability density function, it has a simple MGF given by;

$$M_y(t) = exp\left\{ t\theta + \sigma\left(\sqrt{\alpha^2 - \beta^2} - \sqrt{\alpha^2 - (\beta+t)^2} \right) \right\} \tag{4.2.4}$$

Given the the MGF in equation (4.2.4), it easy to note many properties of NIG that makes it useful to model the randomness of a systematic mortality risk model.

Lemma 4.2.2. *The following are the features of Normal Inverse Gaussian (NIG) distribution that make it the best heavy tailed distributionLillestol (2000).*

(i.) If $Y \sim NIG(\alpha, \beta, \theta, \sigma)$, then $T = kY \sim NIG(\frac{\alpha}{k}, \frac{\beta}{k}, k\theta, k\sigma)$.

(ii.) If $Y_1 \sim NIG(\alpha, \beta, \theta_1, \sigma_1)$ and $Y_2 \sim NIG(\alpha, \beta, \theta_2, \sigma_2)$ are independent, then the sum. $Z = Y_1 + Y_2 \sim NIG(\alpha, \beta, (\theta_1 + \theta_2), (\sigma_1 + \sigma_2))$.

(iii.) If $Y_i \sim NIG(\alpha, \beta, \theta, \sigma), (i = 1, \cdots, m)$ are independent, then the sample mean $\bar{Z} = \frac{1}{m}\sum_{i=1}^{m} Y_i \sim NIG(m\alpha, m\beta, \theta, \sigma)$.

(iv.) If $Y_i \sim NIG(\alpha, \beta, \theta, \sigma)$, then variable $Z = (\frac{Y-\theta}{\sigma})$ has the Standard NIG distribution denoted by $Z \sim NIG(\alpha\sigma, \beta\sigma, 0, 1)$.

Proof. We prove the Lemma 4.2.2 as follows. □

    1. Let $T = kY$. Then

$$M_T(t) = M_{kY}(t) = \mathbb{E}[e^{tkY}] = M_Y(tk)$$

$$= exp\left\{tk\theta + \sigma(\sqrt{\alpha^2 - \beta^2}) - \sqrt{\alpha^2 - (\beta + tk)^2)}\right\}$$

$$= exp\left\{tk\theta + k\sigma(\sqrt{(\tfrac{\alpha}{k})^2 - (\tfrac{\beta}{k})^2)} - \sqrt{(\tfrac{\alpha}{k})^2 - (\tfrac{\beta}{k} + t)^2)}\right\}$$

$$T = kY \sim NIG(\tfrac{\alpha}{k}, \tfrac{\beta}{k}, k\theta, k\sigma)$$

2. Let $Z = Y_1 + Y_2$.

   For identically and independently distributed Y_1, Y_2, we have $M_{Y_1+Y_2}(t) = M_{Y_1}(t)M_{Y_2}(t)$.

   By using the property, we have;

   $$M_Z(t) = exp\left\{tk\theta_1 + \sigma_1(\sqrt{\alpha^2 - \beta^2}) - \sqrt{\alpha^2 - (\beta + tk)^2)}\right\}$$

   $$exp\left\{tk\theta_2 + \sigma_2(\sqrt{\alpha^2 - \beta^2}) - \sqrt{\alpha^2 - (\beta + tk)^2)}\right\}$$

   $$= exp\left\{tk(\theta_1 + \theta_2) + (\sigma_1 + \sigma_2)(\sqrt{\alpha^2 - \beta^2}) - \sqrt{\alpha^2 - (\beta + tk)^2)}\right\}$$

   $$Z = Y_1 + Y_2 \sim NIG(\alpha, \beta, (\theta_1 + \theta_2), (\sigma_1 + \sigma_2))$$

3. Let $\bar{Z} = \sum_{i=1}^{m} Y_i$. From the property of $M_T(t) = M_{\sum_{i=1}^{m} Y_i}(\tfrac{t}{m})$. Since $Y_i, i = 1, 2, 3....m$ and are iid, we get

   $$M_{\sum_{i=1}^{m} Y_i}(\tfrac{t}{m}) = exp\left\{\tfrac{t}{m}\theta + \sigma(\sqrt{\alpha^2 - \beta^2}) - \sqrt{\alpha^2 - (\beta + \tfrac{t}{m})^2)}\right\}$$

   $$= exp\left\{t\theta + \sigma(\sqrt{(m\alpha)^2 - (m\beta)^2} - \sqrt{(m\alpha)^2 - (m\beta + t)^2)}\right\}$$

   $$\bar{Z} = \tfrac{1}{m}\sum_{i=1}^{m} Y_i \sim NIG(m\alpha, m\beta, \theta, \sigma).$$

4. Let $Z = (\tfrac{Y-\theta}{\sigma})$. We use the first property to obtain;

   $$M_Z(t) = \mathbb{E}[exp[(Y-\theta)\tfrac{t}{\sigma}] = \mathbb{E}exp((-\tfrac{Y}{\sigma}t)\times M_Y(\tfrac{t}{\sigma}))$$

   $$= exp(-\tfrac{Y}{\sigma}t) * exp(\theta)\tfrac{t}{\sigma} + \sqrt{(\sigma\alpha)^2 - (\sigma\beta)^2} - \sqrt{(\sigma\alpha)^2 - (\sigma\beta + t)^2}$$

   $$= exp\left\{\sqrt{(\sigma\alpha)^2 - (\sigma\beta)^2} - \sqrt{(\sigma\alpha)^2 - (\sigma\beta + t)^2)}\right\}$$

   $$Z \sim NIG(\alpha\sigma, \beta\sigma, 0, 1)$$

Lemma 4.2.2 proves that *NIG* is the best fit compared to many other commonly used heavy-tailed distributions when modeling SMR under data paucity data structure such as Kenya.

4.2.3 The Parameter Estimations of the NIG Distribution

The sum of NIG variables when used in a model is still a NIG variable with the four parameters. This property may lack in several other common heavy-tailed distributions like student's t distribution, Variance Gamma and Hyperbolic Distributions, which make it sufficient Lillestol (2000) when modeling the mortality risk by the incorporation of the Bühlmann Credibility Approach.

Through solving the mean, variance, kurtosis, and skewness for all of the parameters, we get a closed analytical method of moments solution under a couple of rather fair conditions Karlis (2002). This means we estimate all the NIG distribution parameters as follows;

$$
\begin{cases}
\hat{\theta} = & m - \frac{3s\sqrt{v}}{3k-4s^2-9} \\
\hat{\sigma} = & \frac{3^{1.5}\sqrt{v(k-\frac{5}{3}s^2-3)}}{3k-4s^2-9} \\
\hat{\beta} = & \frac{s}{\sqrt{v(k-\frac{5}{3}s^2-3)}} \\
\hat{\alpha} = & \frac{\sqrt{3k-4s^2-9}}{\sqrt{v(k-\frac{5}{3}s^2-3)}}
\end{cases}
\tag{4.2.5}
$$

where $\hat{\theta} = \hat{\mu}$ and $\hat{\sigma} = \hat{\delta}$ from the estimates.

During the parameter estimations of the NIG distributions, it is important to consider short time horizons, as the mortality of a shorter time or instantaneous rate of death. Thus, rejection of a non-zero mean return null hypothesis is not probable during the study, which contradicts the null hypothesis and makes NIG a suitable distribution for modeling. In addition, it is vital to note that $Y \sim NIG(\hat{\alpha}, \hat{\beta}, \hat{\theta}, \hat{\sigma})$ and all the parameters are estimated from the available Kenyan Mortality data.

4.2.4 Testing For Normality of Mortality Data On Models

In this sub-section, we conduct three important statistical tests to ascertain on whether the Kenyan mortality data exhibit Normality assumptions commonly used in the classical models. The results informs the inclusion of BCA approach into the classical models. The test include;

4.2.4.1 JB Test for Model A

The JB test is used to test whether the kind of Kenya mortality data to be used follows a normal statistical distribution Thadewald and Büning (2007). This test statistic is non-negative, which signals that the mortality-specific data do not have a normal distribution

property Wang, Huang, and Liu (2011).

Definition 4.2.5. Let denote *JB* be a parametric test statistic given by;

$$JB = \frac{m}{12}\left(2S^2 + \frac{1}{2}(K-3)^2\right) \tag{4.2.6}$$

where *m* is the observed data total number, which are degrees of freedom and *S* is defined as sample skewness, while sample kurtosis is denoted as *K* :

$$S = \frac{\hat{\theta}_3}{\hat{\sigma}^3} = \left(\frac{\frac{1}{m}\sum_{i=1}^{m}(x_i - \bar{x})^3}{\left(\frac{1}{m}\sum_{i=1}^{m}(x_i - \bar{x})^2\right)^{3/2}}\right)$$

$$K = \frac{\hat{\theta}_4}{\hat{\sigma}^4} = \left(\frac{\frac{1}{m}\sum_{i=1}^{m}(x_i - \bar{x})^4}{\left(\frac{1}{m}\sum_{i=1}^{m}(x_i - \bar{x})^2\right)^{2}}\right)$$

where $\hat{\theta}_3$ and $\hat{\theta}_4$ are denoted as the estimates of $3rd$ and $4th$ central moments, respectively, $\bar{x}$ is denoted as sample mean where as $\hat{\sigma}^2$ is denoted as the variance estimate.

From the set of Kenyan mortality data, we formulate a null hypothesis against an alternative hypothesis as;

JB($p < 0.05$)= Reject H_0 (A Normal Distribution)

JB($p < 0.05$)= Reject H_0 (Not A Normal Distribution) where p is the p-value.

Distribution	Mean	Standard Deviation
Normal	$(62.565, 64.466)$	$(2.057, 2.455)$
The adjusted test statistic:	$JB =$	$(4.183, 4.258)$
Significance level:	$\alpha = 0.05$	
Critical value:	$(0.1235, 0.1358)$	
Critical region:	Reject H_0 if	$JB > (0.01235, 0.01358)$

Table 4.1: JB Normality test for Model A of Males and Females Respectively

Table 4.1 shows a test statistic and p-value of (4.183, 4.258) and (0.01235, 0.01358) respectively. In the above case, we reject the stated null hypothesis (H_0) and make conclusion that the data is not normally distributed. We conclude at $\alpha = 0.05$ that our Kenyan mortality data is heavy-tailed distributed.

4.2.4.2 DH for Model B

Since model B is a multivariate parameter and model errors using independent univariate Normal Inverse Gaussian Levy processes, Ainou (2011), we used the DH test see Doornik and Hansen (2008) to test whether the Kenyan Mortality data follows a Gaussian distribution before proposing NIG distribution.

We rewrite the equation (3.1.3) in the form;

$$K_{t+1} = K_t + \theta + CZ_{t+1} \tag{4.2.7}$$

where K_t is a two-dimension random walk with drift, θ is the constant, C is a 2×2 upper triangular matrix constant and Z_t is a 2-dimensional standard Gaussian random variable. From equation (4.2.7), we rewrite it as follows;

$$logitq(x,t) = K_{t+1}^1 + K_{t+1}^2(x) \tag{4.2.8}$$

where x starts from ages 60 to 90 and t covers from 2010 to 2020. Then, the linear regression is applied in equation (4.2.7) when estimating the value of K_t. The mean and variance is estimated as follows;

$$\begin{cases} E[K_{t+1} - K_t] = \theta \\ Var[K_{t+1} - K_t]] = CC' \end{cases} \tag{4.2.9}$$

Equation (4.2.9) is showing that the mean as well as the variance of the first consecutive differences, $Z_{t+1} - A_t$, are used when estimating θ and $W = CC'$, respectively. In addition, we do estimations before summarizing our results in the Table 4.2. Commonly, those negative value for θ_1 shows SMR improvement. From the same trend, the positive value for θ_2 indicates that SMR at higher ages, which increases at a slower rate.

$\tilde{\theta}$	$\hat{W}$
−0.0868560	(0.052567300, −0.00066890)
0.00082550	(−0.00018935, 0.0000023457)

Table 4.2: Variance and Mean Matrices Estimates for Model B

The multivariate section of Table 4.3 below has indicated that the test statistic is significant according to the p-value of the test, which is a bi-variate normality assumption that has been rejected at a $\alpha = 0.05$ significance level. Subsequently, the multivariate or bi-variate normality assumption that has been made on the model does not hold.

Normal	Distribution	Kenyan Data
The adjusted test statistic:	$p-value =$	$(24.183, 25.242)$
Significance level:	$\alpha = 0.05$	
Critical value:	$(0.1235, 0.1581)$	
Critical region:	Reject H_0 if	$p-value < (0.01235, 0.01581)$

Table 4.3: DH Normality test for Model B of Males and Females Respectively

We also confirm the DH Normality test for model B using a Multivariate Shapiro-Wilk Test for Normality, which is tabulated in Table 4.4 as;

Normal	Distribution	Kenyan Data
The adjusted test statistic:	$p-value =$	$(0.8629, 0.9025)$
Significance level:	$\alpha = 0.05$	
Critical value:	0.1235	
Critical region:	Reject H_0 if	$p-value < 0.01422$

Table 4.4: A Multivariate Shapiro-Wilk Test for Normality of Males and Females Respectively

From the test statistic of $(0.8629, 0.9025)$ in Table 4.4, which is greater compared to the critical value at the confidence significance level of $\alpha = 0.05$, we then reject the null hypothesis, which assumes that the Kenyan mortality data is normal, which prompts the use of NIG during modeling. In addition, it confirms the findings of the Doornik-Hansen test on the Normality for Model B.

4.2.4.3 The AD Test for Model C

The AD test is a statistical test done when determining whether a specific data sample has been drawn from a particular type of statistical probability distribution see Evans, Drew, and Leemis (2017). We defined our hypothesis as follows;

H_0: Kenyan mortality data follow a Normal distribution vs H_1: Kenya population data do not follow a Normal distribution.

Definition 4.2.6. Let A^2 be the test Statistic given as;

$$A^2 = -M - S \qquad (4.2.10)$$

where $S = \sum_{i=1}^{M} \frac{2i-1}{M} log_e\{G(Y_i) + log_e[1 - [G(Y_{M+1-i})]]\}$ and G is the CDF of the Gaussian distribution and Y_i are the ordered Kenyan Mortality data.

From the analysis;

Distribution	Mean	Standard Deviation
Normal	$(62.004, 64.360)$	$(2.001, 2.816)$
The adjusted test statistic:	$A^2 =$	$(0.8257, 0.8545)$
Significance level:	$\alpha = 0.05$	
Critical value:	$(0.752, 0.788)$	
Critical region:	Reject H_0 if	$A^2 < (0.00752, 0.00788)$

Table 4.5: AD Test for Model C of Males and Females Respectively

From Table 4.5, we deduce that since the test statistic is $(0.8257, 0.8545)$ is less than 0.752 from the AD test, we do not reject the null hypothesis and conclude that we do not have sufficient evidence at $\alpha = 0.05$ to conclude that Kenyan Data is following a Normal distribution (we model it as a NIG distribution).

4.3 Incorporating the BCA into the Mortality Models

We now incorporate these SMR models into BCA;

4.3.1 Mathematical Modeling of Mortality Risk

Definition 4.3.1. Let $p(x,t)$ and $q(x,t)$ denote the likelihood that a person aged exactly x years lives and dies respectively to age exactly $(x+t)$. From the summation of these probabilities of life and death equals to 1. The value of $q(x,t)$ is associated with instantaneous death rates of $\mu(x,t)$ commonly known as force of systematic mortality.

With the assumptions of *UDD* and *CFM* in between the integer age of x and year t, then the force of systematic mortality is equivalent to the central death rates, $\mu(x,t) = m(x,t)$ with the rate of $m(x,t)$ being defined as *CDR*.

From the estimations of parameters to be used in the model, we assume an OLS method since the data given is in a discrete form. We make an assumption of t years within the year of fitting span $\left[x_{low}, x_{high}\right]$ $\left(x_{low} - x_{high} + 1 = t\right)$ and k ages in the fitting age span $\left[t_{low}, x_{high}\right]$ $\left[t_{low} - t_{high} + 1\right] = m$. While the models A and C apply the use of $ln(m(x,t))$, the B model applies $logit(q(x,t)) = ln\left(\frac{q(x,t)}{1-q(x,t)}\right)$ to model rates of mortality. From the empirical Kenyan mortality data, it shows that $ln(m(x,t))$ is in both model A and C and $ln\left(\frac{q(x,t)}{1-q(x,t)}\right)$, which has displayed a downward trend during time period x (see Figure 4.3.1).

As a way of eliminating this downward trend, let us denote $Q(x,t) = ln(m(x,t)) - ln(m(x,t-1))$ for the two models namely A and C models while $Q(x,t) = logit(q(x,t)) - logit(q(x,t-1))$ for the B model (as illustrated in Figure 4.3.2), $x = x_{low} + 1, ..., t_{high}$.

With an assumption of having that $(t-1)$ observed values, the value of $Q_{x,k_{low}+1},\ldots,$ Q_x, k_{high} are illustrated in above Figure (4.3.3).

The BC estimate $\hat{Q}_x, t_{high}+1$ for those ages x in the year $t_{high}+1$ is defined as the weighted average or weighted proportion of the mean sample, $\hat{Q}_x = \frac{1}{m}\sum\limits_{x=x_{low}+1}^{x_{high}} Q_{(x,t)}$, and the exact mean of μ with weights of Z at the same time $(1-Z)$, in that order. The distribution of the parameter risk, $Q_{x,t}$ will always determine the kind of parameters used when determining the value of parameter μ as well as the BC estimate or factor of Z.
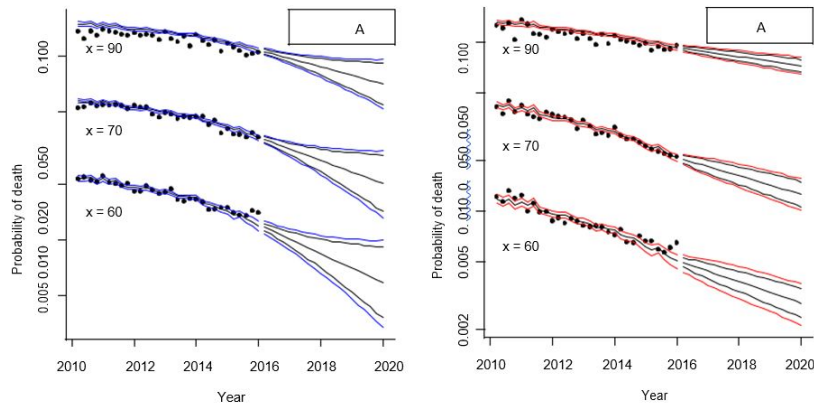
Figure 4.3.1: $log_e\ (m(x,t)\ )$ & $logit(q(x,t)\ )$ against time for Kenyan Males
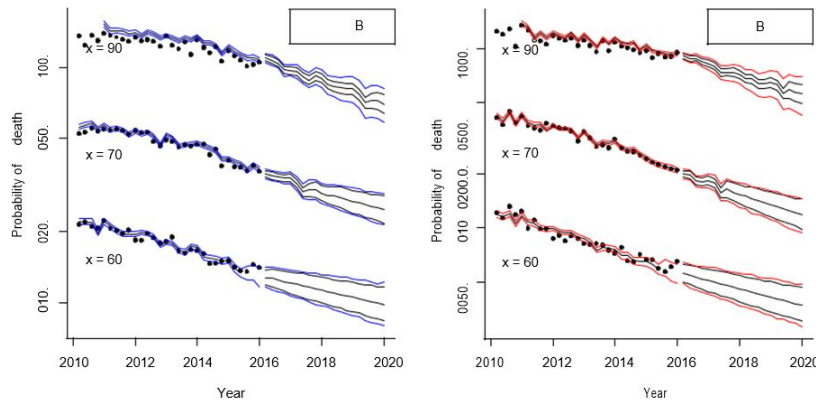
Figure 4.3.2: $log_e\ (m(x,t)\ )$ & $logit(q(x,t)\ )$ against time for Kenyan Females
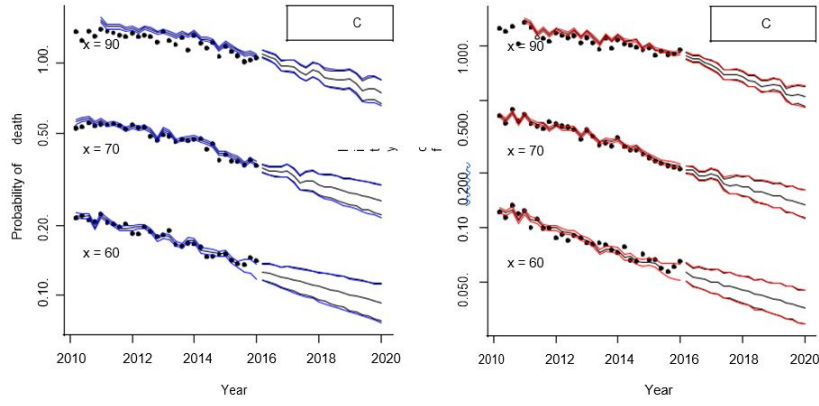
Figure 4.3.3: $Q_{x,t}$ against time for Kenyan Males and Females

4.3.2 Incorporation of the BCA into Model A

Proposition 4.3.1. *Let* $\mu(x,t) = \alpha_x + \beta_x k_t + e(x,t)$ *be under Non-Gaussian assumptions with* $e(x,t)$ *to be following a NIG distribution randomness assumption with mean of* $\theta + \frac{\beta}{\tau}\sigma$ *and variance of* $\sigma\frac{\alpha^2}{\tau^3}$, *then*

$$\begin{bmatrix} e(x,t) \\ e(t) \end{bmatrix} / W_{t-1} \sim NIG \left\{ \begin{array}{cc} \theta + \frac{\beta}{\tau}\sigma, & \sigma_x \frac{\alpha^2}{\tau^3} \\ \theta + \frac{\beta}{\tau}\sigma, & \sigma_\varepsilon \frac{\alpha^2}{\tau^3} \end{array} \right\}$$

where W_{t-1} *is the adapted to filtration about the process up to a time* $t-1$ *and covariances of the two random errors is zero.*

Proof. It is assumed that the overall mortality trend follows a simple random walk with a drift of ϑ for the prediction of mortality such as $\kappa_t = \kappa_{t-1} + \vartheta + e(t)$ where randomness trends of $(e(t))$ of the time follows a *NIG* and are i.i.d such that $e(t)$ for $t = t_x + 1, \ldots.t_m$. Let a random variable $Q_{x,t}$ denote the differences between the central death rates t and $t-1$. This means that $Q_{x,t} = ln(m(x,t)) - ln(m(x,t-1))$. $\qquad\square$

$$Q_{x,t} = \beta_x(k_t - k_{t-1}) + \triangle e(x,t)$$

$$Q_{x,t} = \beta_x\vartheta + \beta_x \times e_t + \triangle e(x,t)$$

where the values of $x = x_1 \ldots, x_m$ and $t = t_{x1}, \ldots, t_{xm}$ and $\triangle e(x,t) \sim NIG(\theta + \frac{\beta}{\tau}\sigma, 2\sigma\frac{\alpha^2}{\tau^3})$ This follows that $Q_{x,t} \sim NIG(\vartheta_1\beta_x, \vartheta_2\beta_x, \beta_x^2\sigma_\varepsilon^2 + 2\sigma_x^2)$, is the sum of independent variables of *NIG* distribution, and still remains a *NIG* distribution with new parameters.

From the conditional expectation as well as variance of $Q_{x,t}$, it is easy to apply the BC, such that $\theta(x) = \mathbb{E}[Q_{x,t}/X] = \beta_x\vartheta$ and $Var[Q_{x,t}/X] = \beta_x^2\sigma_\varepsilon^2 + 2\sigma_x^2$ in the similar

order. Since the expectation of value of the stated hypothetical mean, $\theta = \mathbb{E}[\theta(x)] = \mathbb{E}[\mathbb{E}[Q_{x,t}/X]] = \theta \mathbb{E}[\mathbb{E}[Q_{x,t}/X]]$, the estimated value of θ, denoted by $\hat{\theta}$ is given by;

$$\hat{\theta} = \frac{\hat{\vartheta}}{n} \sum_{x_1}^{x_m} \hat{\theta}_x = \frac{\hat{\vartheta}}{n} \tag{4.3.1}$$

From the equation (4.3.1), the variance process expected value denoted by $a = \mathbb{E}[a(X)] = \mathbb{E}[\beta_x^2]\sigma_\varepsilon^2 + 2\mathbb{E}[\sigma_x^2]$, is estimated as;

$$\hat{a} = \frac{\hat{\sigma_\varepsilon}^2}{n} \sum_{x_1}^{x_m} \hat{\theta}_x^2 + 2 \sum_{x_1}^{x_n} \frac{\hat{\sigma_x}^2}{m} \tag{4.3.2}$$

while the hypothetical mean variance $c = Var[\theta(X)] = \vartheta^2 Var[\beta_X] = \vartheta^2 \mathbb{E}[\beta_X^2] - \mathbb{E}[\beta_X]^2$. This can be estimated through;

$$\hat{a} = \vartheta^2 \left(\frac{\hat{\vartheta}}{n} \sum_{x_1}^{x_m} \theta_x^2 - (\frac{1}{n} \sum_{x_1}^{x_m} \hat{\theta}_x^2) \right)$$

By writing the equation in form of the $Z\overline{X} + (1-Z)\mu$, it is easy to estimate $\hat{\mu}$ as $\frac{\hat{\vartheta}}{n}$.

Hence, the proof.

4.3.3   Incorporation of the BCA into the Model B

Proposition 4.3.2. *Let $\mu(x,t) = k_t^{(1)} + k_t^{(2)}(x - \bar{x}) + e(x,t)$ be under Non-Gaussian assumptions with $e(x,t)$ to be following a NIG distribution randomness assumption with mean of $\theta + \frac{\beta}{\tau}\sigma$ and variance of $\sigma\frac{\alpha^2}{\tau^3}$, then*

$$\begin{bmatrix} e_t^{(1)} \\ e_t^{(2)} \end{bmatrix} / W_{t-1} \sim NIG \left\{ \begin{array}{ccc} \theta_1 + \frac{\beta}{\tau}\sigma_1, & \sigma_{e,1}^2 & \sigma_{e,1}\sigma_{e,2} \\ \theta_2 + \frac{\beta}{\tau}\sigma_2, & \sigma_{e,1}\sigma_{e,2} & \sigma_{e,2}^2 \end{array} \right\}.$$

where W_{t-1} is the adapted to filtration about the process up to a time $t-1$ and covariances of the two random errors is zero.

Proof. Let $k_t^{(1)}$ and $k_t^{(2)}$ be time trends modeled by a bi-variate random walk having a drift of ϑ. It is easy to model $k_t^{(1)}$ and $k_t^{(2)}$ time trends as bi-variate random walk with the drift ϑ i.e. $k_t^{(i)} = k_t^{(i)} + \vartheta + e(x,t)$ where $k_t^{(i)}, i = 1, 2$ to show that $(k_t^{(1)} \ k_t^{(2)})'$, $\vartheta = (\vartheta_1, \vartheta_2)'$ and $e(t) = (e_t^{(1)}, e_t^{(2)})'$.

The two errors $e(x,t)$ as well as $e(t)$ assumed to be i.i.d. while the time trends for $\forall$ values of t are also *i.i.d.* when values of $i = 1, 2$ and W_{t-1} is the adapted to filtration about the process up to a time $t-1$.

By considering a random variable;

$Q_{x,t} = ln\left(\frac{q(x,t)*p(x,t-1)}{p(x,t)*q(x,t-1)}\right) = logit(p(x,t) - logit(p(x,t-1)$

$Q_{x,t} = (k_t^{(1)} - k_{t-1}^{(1)}) + (x - \bar{x})(k_t^{(2)} - k_{t-1}^{(2)}) + \Delta e(x,t)$

$Q_{x,t} = \{(\vartheta_1 + \vartheta_2)(x - \bar{x})\} + \{(e(t) + e(t-1)) * (x - \bar{x})\} + \Delta e(x,t)$

for values of $x = 1, 2, 3....., n$, $t = 1, 2, 3, ...m$, and $\Delta e(x,t) = e(x,t) - e(x) \sim NIG\left(0, 2\sigma_x^2\right)$.

$Q_{x,t} \sim NIG(\vartheta_1 + \vartheta_2)(x - \bar{x})\}, (x - \bar{x})\sigma_{e,1}^2 + 2*(x - \bar{x})\sigma_{e,1}\sigma_{e,2} + 2\sigma_x^2)$.

The conditional expectation variance of $[Q_{x,t}/X] = (\vartheta_1 + \vartheta_2)(x - \bar{x})$ and the variance expectation $\text{Var}(\upsilon) = \hat{\upsilon} = (x - \bar{x})\sigma_{e,1}^2 + 2(x - \bar{x})\sigma_{e,1}\sigma_{e,2} + 2\sigma_x^2$.

It is essential to estimate hypothetical mean expected value, $\theta = \mathbb{E}[\theta(X)] = \vartheta_1 + \vartheta_2 1 + \mathbb{E}(x - \bar{x})$, with its estimation as:

$$\hat{\theta} = \hat{\vartheta}_1 + \hat{\vartheta}_2 + \frac{1}{n}\sum_{x_1}^{x_m}(x - \bar{x})$$

with $\hat{\theta} = \frac{1}{n(m-1)}\left(\sum_{x_1}^{x_m}[ln(m(x,t_1 - 1)) - ln(m(x,t_1 - 1))]\right) = \frac{1}{n}\sum_{x_1}^{x_m}\hat{Q}_{x,t} = \bar{Q}_{x,t}$.

Similarly, the process variance expected value, $Var(\upsilon) = \hat{\upsilon} = (x - \bar{x})\sigma_{e,1}^2 + 2(x - \bar{x})\sigma_{e,1}\sigma_{e,2} + 2\sigma_x^2$ is estimated as;

$$Var(\upsilon) = \hat{\sigma}_{e,1}^2 + \frac{\sigma_{e,2}^2}{n}\sum_{x_1}^{x_m}(x - \bar{x})^2 + \frac{1}{n}\sum_{x_1}^{x_m}2\hat{\sigma}_x^2$$

The expected value of $\hat{\upsilon}$ is estimated as;

$$\hat{\upsilon} = \hat{\vartheta}_1^2\frac{1}{n}\sum_{x_1}^{x_m}(x - \bar{x})^2$$

since $Var[\theta(\vartheta)] = \vartheta_2^2 Var[(x - \bar{x})]$. Thus, for values of ϑ and $\hat{v}$, w have $Z = \left(\frac{n}{n + \frac{\vartheta}{\hat{v}}}\right)$.

Hence, the proof. □

4.3.4   Incorporation of the BCA into the Model C

Proposition 4.3.3. *Let $\mu(x,t) = ln(m(x,t)) = \kappa_t^{(0)} + \kappa_t^{(1)}ln(m(x,t_x - 1)) + e(x,t)$ be under Non-Gaussian assumptions with $e(x,t)$ to be following a NIG distribution randomness assumption with mean of $\theta + \frac{\beta}{\tau}\sigma$ and variance of $\sigma\frac{\alpha^2}{\tau^3}$, then*

$$\begin{bmatrix} e(x,t) \\ e(0,t) \\ e(1,t) \end{bmatrix} / W_{t-1} \sim NIG \left\{ \begin{array}{l} \theta_1 + \frac{\beta}{\tau}\sigma \\ \theta_2 + \frac{\beta}{\tau}\sigma \quad, \sigma^2, \sigma_{\varepsilon,0}^2, \sigma_{\varepsilon,1}^2 \\ \theta_3 + \frac{\beta}{\tau}\sigma \end{array} \right\}$$

where W_{t-1} is the adapted to filtration about the process up to a time $t - 1$.

Proof. The time trend in the model $\hat{k}_t^{(j)}$, $j = 0, 1$, is assumed to be following a simple random walk of drift ϑ_i, i.e., $\hat{k}_t^{(j)} = \hat{k}_{t-1}^{(j)} + \vartheta_i + e(j,t)$ since $j = 0, 1$ and the term of $e(x,t)$ is denoted as the error term associated with the model. $e_{(x,t)}$ is assumed to be *NIG* distributed at the same time i.i.d. with mean of $\theta + \frac{\beta}{\tau}\sigma$ and variance of $\sigma_x\frac{\alpha^2}{\tau^3}$ for $\forall$ values of t.

By considering the random variable given as $Q_{x,t}$ defined by $ln(m(x,t)) - ln(m(x,t-1))$ and replacing it with the values of the model as written in equation (3.1.5), we obtain;   □

$$Q_{x,t} = k_t^{(0)} + k_{t-1}^{(0)} + k_t^{(1)} + k_{t-1}^{(1)} + ln(m(x,t_1 - 1) + \Delta e(x,t)$$

$$Q_{x,t} = \vartheta_0 + \vartheta_1 ln(m(x,t_1 - 1) + [\varepsilon(0,t) + \varepsilon(0,1)ln(m(x,t_1 - 1)] + \Delta e(x,t)$$

where $x = 1,2,3.....,n$, $t = 1,2,3,.....m$, and $\Delta e(x,t) = e(x,t) - e(x,t-1) \sim NIG(0,2\sigma^2)$ and from equation (4.3.7), we have;

$$Q_{x,t} \sim NIG(\vartheta_0 + \vartheta_1 ln(m(x,t_1 - 1), \sigma^2(\varepsilon,0) + [ln(m(x,t_1 - 1))]^2 \sigma^2(\varepsilon,1) + 2\sigma^2] \quad (4.3.3)$$

At the specific ages, we make an assumption that $\Delta e(x,t)$ follows a White Noise in a way that $Q_{x,t}$ are identically and independently distributed for the values of $x = 1,2,3.....,n$. From the above model, we apply the approach of BC theory and calculate the values of the hypothetical mean and variance, where;

$$\theta(X) = \mathbb{E}(Q_{X,t}/X) = \vartheta_0 + \vartheta_1 ln(m(x,t_1 - 1)$$

$$\text{Var}(\upsilon) = Var(Q_{X,t}/X) = (\sigma^2(\varepsilon,0) + [ln(m(x,t_1 - 1)]^2 \sigma^2(\varepsilon,1) + 2\sigma^2)$$

We proceed ahead with the estimation of the values of the means and variance, $\theta(X)$ and Var(υ) respectively as follows;

$$\hat{\theta} = \hat{\vartheta}_0 + \left[\frac{\hat{\vartheta}_1}{n} \sum_{x_1}^{x_n} ln(m(x,t_1 - 1)) \right] \quad (4.3.4)$$

From equation (3.1.6), it important to rewrite equation (4.3.4) in the BC formulae:

$$\bar{Q}_{X,t} = \left(\frac{(ln(m(x,t_1 - 1) - ln(m(x,t_1 - 1))}{m-1} \right) = \frac{1}{m-1} \sum_{x_1}^{x_n} (Q_{X,t})$$

Consequently, we estimate the expected variance, Var(υ) as used in the BCA as; Var(υ) =Var($Q_{X,t}/X) = (\sigma^2(\varepsilon,0) + [ln(m(x,t_1 - 1))]^2 \sigma^2(\varepsilon,1) + 2\sigma^2)$ as;

$$\text{Var}(\upsilon) = \hat{\upsilon} = \hat{\sigma}^2(\varepsilon,0) + \frac{\hat{\sigma}^2(\varepsilon,1)}{n} \sum_{x_1}^{x_n} [ln(m(x,t_1 - 1))]^2 + 2\hat{\sigma}^2$$

while the variance of the stated hypothetical mean, $\upsilon = Var[\theta(X)] = Var[ln(m(x,t_1 - 1))] = \vartheta_1^2 \mathbb{E}[ln(m(x,t_1 - 1)^2] - \mathbb{E}[ln(m(x,t_1 - 1)]^2$.

Thus, for values of ϑ and $\widehat{v}$, w have $Z = \left(\frac{n}{n + \frac{\vartheta}{\widehat{v}}} \right)$.

Hence, the proof.

Remark 4.3.1. Using the probability theory for calculation of the variance is estimated as;

$$Var(\upsilon) = \hat{\upsilon} = \hat{\vartheta}_1^2 \frac{1}{n} \sum_{x_1}^{x_m} ln(m(x,t_1 - 1))^2 - \left[\frac{1}{n} \sum_{x_1}^{x_m} ln(m(x,t_1 - 1)) \right]^2 \} \quad (4.3.5)$$

where estimation of the parameters, $\hat{\upsilon}$, $\hat{\vartheta}$ and $\hat{\theta}$ as the the BC estimates of the $\bar{Q}_{X,t}$. In addition, this has a value of;

$$\hat{\bar{Q}}_{(X,t+1)} = \begin{cases} ln\left(\hat{m}_{x,t_n+1}\right) - ln\left(m_{x,t_n+1}\right) & , \quad for\ A/C \\ logit\left(\hat{q}_{x,t_n+1}\right) - logit\left(q_{x,t_n+1}\right) & , \quad for\ B \end{cases} = Z\hat{Q}_x + (1-Z)\hat{\theta}$$

(4.3.6)

where Z is the BC estimates of $logit\left(q_{x,t_n+1}\right)$ and $ln(m_{x,t_n+1})$ for a given age x for a period of t_n+1. Hence the estimates are;

$$\begin{cases} ln(\hat{m}_{x,t_n+1}) = ln(m_{x,t_n}) + Z\hat{Q}_x + (1-Z)\hat{\theta} \\ logit(\hat{q}_{x,t_n+1}) = logit(q_{x,t_n}) + Z\hat{Q}_x + (1-Z)\hat{\theta} \end{cases}$$

(4.3.7)

It is vital to note that Propositions **4.3.1-4.3.3**. conforms to the models defined in Definitions **3.1.1-3.1.3**, where we estimate four parameters of NIG distribution as opposed to the two parameters of a Normal distribution. In addition, the randomness is no longer following a Gaussian distribution with mean of zero and variance of σ^2 as often used in the classical Lee and Carter (1992), Cairns, Blake, and Dowd (2006) and Tsai and Yang (2015a) models.

4.3.5 BC Estimate Determination

When dealing with the three models namely A, B, and C respectively, the parameters under BCA are determined under MLE method as tabulated in the Table 4.6;

Estimation	for values of		θ, υ and b
$\bar{Q}_{x_i} = (Q_{x_i,t+1}, ... Q_{x_i,t+n})$	$\bar{Q}_{x_i} = \frac{1}{m-1}\sum\limits_{x_i=1}^{x_n} Q_{x_i+t}$	$\hat{v}_{x_i} = \frac{1}{m-2}\sum\limits_{x_i=1}^{x_n} \left(Q_{x_i+t} - \bar{Q}_{x_i+t}\right)^2$	
...	.....	...	
$\bar{Q}_{x_N} = (Q_{x_i,t+1}, ... Q_{x_i,t+N})$	$\bar{Q}_{x_N} = \frac{1}{m-1}\sum\limits_{x_i=1}^{x_N} Q_{x_i+t}$	$\hat{v}_{x_N} = \frac{1}{m-2}\sum\limits_{x_i=1}^{x_N} \left(Q_{x_i+t} - \bar{Q}_{x_i+t}\right)^2$	
$\hat{b} = \frac{1}{n-1}\sum\limits_{x_i=1}^{x_n} \left(\bar{Q}_{x_i} - \bar{Q}\right)^2 - \frac{\hat{v}}{m-1}$	$\hat{\theta} = \bar{Q} = \frac{1}{m}\sum\limits_{x_i=1}^{x_n} \bar{Q}_x$	$\hat{v} = \frac{1}{n}\sum\limits_{x_i=1}^{x_n} \hat{v}_x$	

Table 4.6: Estimations for the values of θ, υ and b

Remark 4.3.2. It is essential to note that the values are as follows $\mathbb{E}[Var[X|\Theta]] = \theta$ and $Var[\mathbb{E}[X|\Theta]] = b$ respectively from the above table. It is vital to take note that values of $\hat{b}$ could sometimes be negative because of subtraction. Whenever it happens, it is essential to customize it by setting the value of $\hat{b}$, which implies that the value of $Z = 0$, and thus the value of BC Estimate becomes $\hat{\theta} = \bar{Q}$.

For the fitting as well as forecasting purposes of the respective three SMR models namely; A, B, and C, we use the following strategies;

4.3.5.1 Strategy EW: expansion of the Window by a year

As a way of predicting the estimates of BC for the year, $t_n + 1$, we apply the estimate $\{\hat{Q}_{X,t_n+1}$ to $Q_{x,t_0+1}, Q_{x,t_1+1}, \ldots Q_{x,t_n+N}\}$, then the value of estimate at time $t_n + 2$, will be obtained by the value of $\overline{Q}_{X,t_n+1} = \frac{1}{m} \left[\sum_{x_1}^{x_n} Q_{X,t} + \hat{Q}_{x,t_1+1} \right]$, $\hat{\theta}(t_n+2) = \frac{1}{n} \sum_{x_1}^{x_n} \hat{Q}_{X,t_n+2}$ and $Z(t_n + 2) = \frac{n}{n+K}$. To estimate the BC estimate for the different time lags i.e. $\tau = 2, 3 \ldots$, we apply the following equation;

$$\bar{Q}_{X,t+\tau} = \frac{1}{m+\tau-2} \left(\sum_{x_1}^{x_n} Q_{X,t} + \sum_{x_1+1}^{x_n-\tau+1} \hat{Q}_{X,t} \right) \tag{4.3.8}$$

where the values of $\hat{\theta}(t_n+2) = \hat{Q}_{x,t_1+\tau}$ and $Z(t_n+\tau) = \frac{n+\tau-2}{n+\tau-2+K}$ and $K = \frac{\mathbb{E}[Var[X|\Theta]]}{Var[\mathbb{E}[X|\Theta]]}$. for those values of for $\bar{Q}_{X,t+\tau}$, $\tau = 2, 3, 4 \ldots$. Therefore, $Z(t_n+\tau)$ is increasing in the value of τ for this strategy of EW.

4.3.5.2 Strategy MW: Movement of the Window by a year

To predict the estimates of BC for the year, $t_n + 1$, we will apply the estimate $\{\hat{Q}_{X,t_n+1}$ to $Q_{x,t_0+1}, Q_{x,t_1+1}, \ldots Q_{x,t_n+N}\}$, then the value of estimate at time $t_n + 2$, will be obtained by the value of $\bar{Q}_{X,t_n+1} = \frac{1}{m} [\sum_{x_1}^{x_n} Q_{X,t} + \hat{Q}_{x,t_1+1}]$, $\hat{\theta}(t_n+2) = \frac{1}{n} \sum_{x_1}^{x_n} \bar{Q}_{X,t_n+2}$ and $Z(t_n + 2) = \frac{n}{n+K}$. To estimate the BC estimate for the different time lags i.e. $\tau = 2, 3 \ldots$, we apply the following equation;

$$\bar{Q}_{X,t+\tau} = \frac{1}{m-1} \left(\sum_{x_1+1}^{x_n-\tau+1} \hat{Q}_{X,t} \right) \tag{4.3.9}$$

and the value of

$$Z(t_n + \tau) = \left(\frac{n-1}{n-1+\hat{K}} \right) \tag{4.3.10}$$

where the values of $\hat{\theta}(t_n+2) = \hat{Q}_{x,t_1+\tau}$ and and $\hat{K} = \frac{\mathbb{E}[Var[X|\Theta]]}{Var[\mathbb{E}[X|\Theta]]}$. for those values of for $\bar{Q}_{X,t+\tau}$, $\tau = 2, 3, 4 \ldots$. Therefore, $Z(t_n+\tau)$ is increasing in the value of τ for this MW strategy.

For all the two strategies, we can get the values of $Z(t_n + \tau)$ where $\hat{Q}_{X,t} = ln(\hat{m}(x,t) - ln(\hat{m}(x,t-1)$ for the A and C models while for B model, $\hat{Q}_{X,t} = ln(\hat{q}(x,t) - ln(\hat{q}(x,t-1)$; all for the values of $(t = t_1 + 1, t_1 + 2, \ldots, t_1 + \tau + 3)$.

This means that the forecasted mortality rates for person aged exactly x in year $t_{low} + \tau$ under the A, B and C models without credibility incorporated approach are all linear functions of τ with given different gradients.

Model **A** is defined as:

$$ln\left(\hat{m}(x,t_n+\tau)\right) = ln\left(\hat{m}(x,t_n)\right) + \hat{\vartheta}\hat{\beta}_x \times \tau$$

$$ln\left(\hat{m}(x,t_n+\tau)\right) = ln\left(\hat{m}(x,t_n)\right) + \hat{Q}^A_{X,t_n+1} \times \tau \tag{4.3.11}$$

Model **B** is defined as;

$$ln\left(\hat{q}(x,t_n+\tau)\right) = ln\left(\hat{q}(x,t_n)\right) + \left[\hat{\vartheta}_1 + \hat{\vartheta}_2\left(x-\bar{x}\right)\right] \times \tau$$

$$ln\left(\hat{q}(x,t_n+\tau)\right) = ln\left(\hat{q}(x,t_n)\right) + \hat{Q}^B_{X,t_n+1} \times \tau \tag{4.3.12}$$

Model **C** is defined as;

$$ln\left(\hat{m}(x,t_n+\tau)\right) = ln\left(\hat{m}(x,t_n)\right) + \left[(\hat{\vartheta}_1 + \hat{\vartheta}_1)ln\left(\hat{m}(x,t_n-1)\right)\right] \times \tau$$

$$ln\left(\hat{m}(x,t_n+\tau)\right) = ln\left(\hat{m}(x,t_n)\right) + \hat{Q}^C_{X,t_n+1} \times \tau \tag{4.3.13}$$

Using the above equations (4.3.11, 4.3.12, and 4.3.13), we get the predicted mortality rates under the three models; A, B, and C. In addition, it is expected that the EW strategy would show the downward trends of all estimated future mortality rates better for each of all ages x. Due to the two common invariant properties of the MW strategy, it is easier to compute the estimates for BC before comparing the EW strategy.

4.3.6   Parameter Estimations for Systematic Mortality Risk Models

1. Model A

On model A, it is important to note that the parameters are subjected to two constraints namely $\sum_{x=1}^{n} \beta_x = 1$ and $\sum_{t=1}^{n} k_t = 0$ as well as the estimations using the method of singular value decomposition (SVD). From the constraint that $\sum_{t=1}^{n} k_t = 0$, it is key to note that the parameter of a_x denoted as $\hat{a}_x$ can be estimated as;

$\hat{a}_x = \frac{1}{n} \sum_{t=1}^{n} ln(m(x,t))$ for values of $t = 1,2,......n$. In addition, the constraint of $\sum_{x=1}^{n} \beta_x = 1$ will lead to the estimates of k_t, which is $\hat{k}_t$ as follows;

$$\hat{k}_t = \sum_{t=1}^{n} \left[ln(m(x,t)) - \hat{a}_x\right]$$

for values of $t = 1,2,......n$. The value of $\hat{\beta}_x$ is obtained through the process of regression of the $[ln(m(x,t)) - \hat{a}_x]$ on the value of $\hat{k}_t$ without involving the constant term being included in all ages of x.

The drift parameter, ϑ, the variance of the time trend error, σ_e^2, and the variance of the model error, σ_x^2, are estimated by the process of;

$$\begin{cases} \hat{\vartheta} = & \frac{1}{n-1} \sum_{t=1}^{n} (k_t - k_{t-1}) = \frac{(k_t - k_1)}{n-1} \\ \sigma_e^2 = & \frac{1}{n} \sum_{t=1}^{n} (k_t - k_{t-1} - \hat{\vartheta})^2 \\ \sigma_x^2 = & \frac{1}{n} \sum_{t=0}^{n} [ln(m(x,t)) - \hat{\alpha}_x + \hat{\beta}_x \hat{k}_t]^2 \end{cases} \quad (4.3.14)$$

The logarithm of the predicted central death rates for all ages of x in year $t + \eta$ is given by

$ln(\hat{m}(x,t+\eta)) = \hat{\alpha}_x + \hat{\beta}_x(\hat{k}_t + \eta(\hat{\vartheta})) = ln(\hat{m}(x,t)) + \hat{\beta}_x(\eta)\hat{\vartheta}$ for all values of $\eta = 1, 2, 3, ...$

2. Model B

For model B, the values of $k_t^{(1)}$ and $k_t^{(2)}$ are determined by the $logitq(x,t)$ on the value of $(x - \bar{x})$ at each value of t that satisfies the following condition;

$$\hat{k}_t^{(1)} = \frac{1}{n} \sum_{t=1}^{n} [logitq(x,t) - \frac{\hat{k}_t^{(2)}}{n} \sum_{t=1}^{n} (x - \bar{x})$$

$$\hat{k}_t^{(1)} = \frac{1}{n} \sum_{t=1}^{n} logitq(x,t)$$

The drift parameter, ϑ, the variance of the time trend error, $\sigma_{ei}^2, i = 1, 2$, and the variance of the model error, σ_x^2, are estimated by the process of;

$$\begin{cases} \hat{\vartheta} = & \frac{1}{n-1} \sum_{t=1}^{n} (k_t^{(i)} - k_{t-1}^{(i)}) = \frac{(k_t^{(i)} - k_1^{(i)})}{n-1} \\ \sigma_e^2 = & \frac{1}{n} \sum_{t=1}^{n} (k_t^{(i)} - k_{t-1}^{(i)} - \hat{\vartheta}_i)^2 \\ \sigma_x^2 = & \frac{1}{n} \sum_{t=0}^{n} [ln(m(x,t)) - k_t^{(1)} - k_{t-1}^{(2)}(x - \bar{x})]^2 \end{cases} \quad (4.3.15)$$

The *logit* function of the predicted mortality rate $logit\hat{q}(x,t)$ for all ages x in year $t + \eta$ is given by $logit\hat{q}(x,t) = k_t^{(1)} + \eta\hat{\vartheta}_1 + k_t^{(2)} + \eta\hat{\vartheta}_2(x - \bar{x}) = logit\hat{q}(x,t) + \hat{\vartheta}_1 + \hat{\vartheta}_2(x - \bar{x})\eta$ for all values of $\eta = 1, 2, 3, ...$

3. Model C

The drift parameter, $\vartheta_i, i = 0, 1$, the variance of the time trend error, $\sigma_{ei}^2, i = 1, 2$, and the variance of the model error, σ_x^2, are estimated by the process of;

$$\begin{cases} \hat{\vartheta}_i = & \frac{1}{n-1} \sum_{t=1}^{n} (k_t^{(i)} - k_{t-1}^{(i)}) = \frac{(k_t^{(i)} - k_1^{(i)})}{n-1} \\ \sigma_{e,i}^2 = & \frac{1}{n} \sum_{t=1}^{n} (k_t^{(i)} - k_{t-1}^{(i)} - \hat{\vartheta}_i)^2 \\ \sigma^2 = & \frac{1}{n(m-1)} \sum_{t=0}^{n} \sum_{x=0}^{m} [ln(m(x,t)) - k_t^{(1)} - k_{t-1}^{(2)} ln(m(x,t-1))]^2 \end{cases} \quad (4.3.16)$$

58

The logarithm of the predicted central death rates for all ages of x in year $t + \eta$ is given by
$ln(\hat{m}(x, t + \eta)) = k_t^{(1)} + \eta \hat{\vartheta}_1 + k_t^{(2)} + \eta \hat{\vartheta}_2 = ln(\hat{m}(x,t)) + [\hat{\vartheta}_1 + \hat{\vartheta}_2(\eta) ln(\hat{m}(x, t - 1)] $ for all values of $\eta = 1, 2, 3, ...$

The equations 4.3.14-4.3.16 shows the parameter estimations for Systematic Mortality Risk Models necessary for modelling under BCA.

4.4 Fitting and forecasting of Models

In this part of the study, we did fitting of the three mortality models, namely A, B, and C, with as well as without incorporation of credibility before making sample-based predictions for future consecutive years using Kenyan mortality data. Given that we do our study for a period of $[T_1, T_2]$ we assume that we end at year t_n during SMR projections. After the projections, we do evaluation of the forecasting performances for the respective years $t_n + 1, ..., T_N$ through application of the mortality data within the rectangle (window) that is defined by $[x_n, x_N] \times [t_n, t_N]$ where $T_1 \leq t_n$ and $t_N < T_2$.

We also do an examination of the forecasting performances in two cases namely before and after incorporation of the BC method. To measure the forecasting error of the true SMR (q) and predicted one $(\hat{q})$, we apply the *MAPE* and *RMSE* measures.

Definition 4.4.1. Let define *MAPE* and *RMSE* for a life aged x in year $t_N + t$ for given a specified fitting year span $[t_n, t_N]$ as;

$$RMSE_{x, t_n + 1} = \sqrt{\left\{ \frac{[\hat{q}(x, t_n + 1) - q(x, t_n + 1)]^2}{T_N - t_n} \right\}} \qquad (4.4.1)$$

The values of the tabulated central death rates for Males and Females were determined by dividing all the yearly observations of the age-specific death numbers by matching the exposed number of population sizes to the death risk of death for the specific period.

For the given sets of age span before the fitting yearly span $[t_n, t_N]$, we will forecast the SMR for the year span under a mortality model. This should help in calculation of the estimates of BC method for each of the years $t = 1, 2, 3....N$ with the application of EW as well as MW strategies.

Definition 4.4.2. We measure the predicting performances by calculation of the average of the *MAPE* over the ages $25, ..., 100$ and predicting years for the remainder of the years. The *AMAPE* will be defined as;

$$AMAPE_{x, t_n + 1} = \frac{1}{T_N - t_n} \sum_{x_1 + 1}^{x_n - \tau + 1} \left(\sum_{x=25}^{100} |\frac{\hat{q}(x, t_n + 1) - q(x, t_n + 1)}{q(x, t_n + 1)}| \right) \qquad (4.4.2)$$

The numerical illustrations prove that incorporating the BCA into these mortality models, A, B, and C, helps improve forecasting performances.

Model **A** with and without BC for Kenyan Males and Females Table is shown in Table (4.7);

Kenya	AMAPE			RMSE	
	Model %	**EW%**	**MW%**	**EW%**	**MW%**
2010	14.85(12.63)	8.56(7.25)	9.55(8.95)	123.50(118.26)	182.20(180.18)
2020	16.45(14.85)	8.23(7.92)	10.16(9.14)	126.45(122.34)	193.47(188.22)
2030	18.86(15.24)	9.34(8.35)	10.89(9.85)	132.75(126.36)	195.89(190.26)
Average	**16.72(14.24)**	**8.71(7.84)**	**10.20(9.31)**	**127.56(122.32)**	**190.52(186.22)**

Table 4.7: Model **A** With and Without BC for Males and Females(brackets)

Model **B** with and without BC for Kenyan Males and Females Table is shown in Table (4.8);

Kenya	AMAPE			RMSE	
	Model %	**EW%**	**MW%**	**EW%**	**MW%**
2010	11.45(14.33)	7.85(8.11)	9.55(9.22)	122.33(115.55)	202.26(195.05)
2020	12.05(15.25)	8.25(8.44)	10.15(9.45)	128.36(123.45)	200.53(198.85)
2030	13.25(15.83)	9.45(9.45)	11.23(9.90)	134.37(127.10)	201.55(196.89)
Average	**12.25(15.14)**	**8.52(8.67)**	**10.31(9.52)**	**128.35(122.03)**	**201.45(196.93)**

Table 4.8: Model B with and without BC for Males and Females(brackets)

Model **C** with and without BC for Kenyan Males and Females Table is shown in Table (4.9);

Kenya	AMAPE			RMSE	
	Model %	**EW%**	**MW%**	**EW%**	**MW%**
2010	15.95(12.44)	7.96(7.35)	8.65(9.15)	140.33(132.52)	210.15(201.22)
2020	16.20(12.90)	8.15(8.55)	8.96(9.35)	141.65(134.35)	209.45(199.35)
2030	17.28(13.35)	9.06(8.95)	9.45(9.88)	139.45(136.36)	211.35(199.95)
Average	**16.48(12.90)**	**8.39(8.28)**	**9.02(9.46)**	**140.48(134.42)**	**210.32(203.17)**

Table 4.9: Model C with and without BC for Males and Females(brackets)

The numerical demonstrations in tables 4.7-4.9 show that incorporating the BC method into the model A, B, and C, outstandingly boost their forecasting performances; however, using the two strategies has contributed to similar prediction performances. Secondly, incorporating the BC does drive the forecasting MAPE ratios from the above three SMR models, thus converging to a consistent level and lower level.

4.5 Analysis and Results

In this chapter, we have incorporated the concept of BC to boost the forecasting accuracy and performances of A, B, and C mortality models. It is crucial to note that error terms of the three classical models are modelled as a NIG distribution as a proposition as statistical tests tabulated in Table (4.1-4.5), showing the non-Gaussian property of Kenyan population data.

With Kenyan mortality data for both males and females, the chapter looked at the fitting of the models as shown in Figures (4.3.1–4.3.3) and shows a perfect fit. In addition, they incorporated the BCA under the common EW as well as MW strategies, thus contributing much better and accurate forecasting performances from lower rates of MAPE and RMSE from Table (4.7-4.12). Therefore, reducing the commonly experienced valuation errors with many other life assurance products.

Chapter 5

Systematic Mortality Risk Forecasting Under Deep Learning Technique

This chapter incorporated the novel DL techniques concept in systematic mortality risk modeling to increase their predictability and forecasting accuracy. It is by applying the Long-Short Term Memory structure (an artificial LSTM architecture) compared to traditional statistical ARIMA (p, d, q) models. The novel deep learning approach helped integrate the CBD model to enhance its accuracy and predictive capacity for future systematic mortality risk in countries with limited data availability, such as Kenya. The results showed that Long Short-Term Memory network architecture had higher levels of precision when predicting the future systematic mortality risks than traditional statistical methods.

5.1 Deep Learning Integration

Definition 5.1.1. Let Neural Networks within the deep learning infrastructure typically consist of sets of input units with multiple hidden layers, which means that more such layers imply a deeper network with hidden units known as nodes or neurons.

In Figure 5.1.1, a neuron is represented in every node connected from one to the other using arcs representing all synapses. Additionally, the graph represents the general input, latent, as well as output variables. A schematical view of an artificial neural network (ANN) has circles representing neurons with lines representing synapses. A Neural networks infrastructure can be represented in Figure 5.1.1;
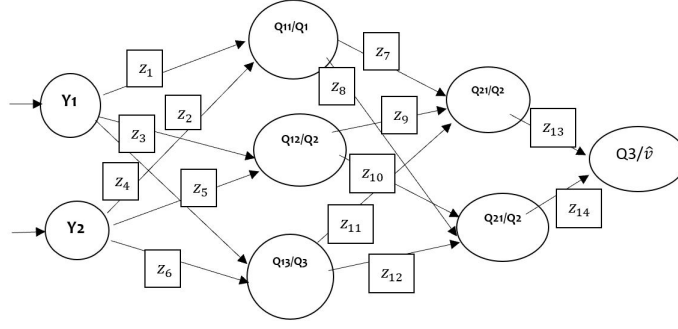
Figure 5.1.1: A Normal Representation of feed-forward Artificial Neural Network (ANN)

The synapses take the individual inputs before multiplying them by a "weight," commonly known as input "strength," to determine the general output. In addition, Neurons are added to these outputs from all available synapses before applying the activation function.

Definition 5.1.2. Let Q be a single neuron called *perceptron* defined by;

$$Q = \Theta(Z^T y + c) \tag{5.1.1}$$

where $y \in \mathbb{R}^\Psi$ is the input and $Z \in R^\Psi$ is the connected synaptic weight, $\psi \in N$ are numbers of the input signals and Θ is the activation function. We represent this term c as the bias associated with the model known as activation verge or threshold. The user must note that the function, Θ, should have a differential because the learning equations have gradients.

One can use The Multilayer Perceptron (MLP) introduction in nonlinear separable problems such as XOR since ANN with a single layer is always not appropriate, thus solving the stated problem. In addition, most neurons in MLP are predisposed on a wide variety of layers, with every unit fully connected to those of the preceding layer, as illustrated by Skovajsová (2017).

The synapses connect units by defining different types of available networks in the system. In an ANN classical pattern like feed-forward ANN, the information moves in a unilateral direction from an input to an output layer simultaneously the Recurrent Neural Networks. RNN processes the information cyclically by using the extra synapses, ensuring that the reprocessed output results are from the entire elaboration process.

Definition 5.1.3. Let $Q \in \mathbb{R}^{k_h}$ be the *ReLU* output functions of the deep neural networks of a generic hidden layer having k_h neurons be

$$Q_1 = \Theta\left(Z^T y + c\right) \tag{5.1.2}$$

where $Z \in R^{\Psi * k_h}$ is defined as a weight matrix and $c \in R^{k_h}$ is called the biases vector.

According to *MLP* scheme, the hidden layer output becomes the input instrument for the following layer.

Lemma 5.1.1. *Let us consider a given problem of regression, where $f \in \mathbb{N}$ defined as the number hidden layersBengio, Goodfellow, and Courville (2017), then the output of $\hat{y} \in \mathbb{R}$ can be calculated by*

$$Q_1 = \Theta_1 \left(Z_1^T y + c_1 \right)$$

$$Q_2 = \Theta_2 \left(Z_2^T Q_1 + c_2 \right)$$

$$Q_3 = \Theta_3 \left(Z_3^T Q_2 + c_3 \right)$$

$$........$$

$$Q_f = \hat{y} = \Theta_f \left(Z_f^T Q_{f-1} + c_f \right)$$

where $Z_1, Z_2, Z_3, ... Z_f$ denote weight matrix vectors, $c_1, c_2, c_3, ... c_f$ denote bias vectors, and $\Phi_1, \Phi_2, \Phi_3, ... \Phi_f$ denote activation functions that needs not be different from one another.

It is vital to note that all measurements of the weight matrices as well as bias vectors do rely on the unit number within the hidden layers; hence, by enhancing these hidden layers in numbers, the abstraction levels of the input data also increase significantly.

5.2 SMR Modeling Under Deep Learning

5.2.1 Mathematical Structure of Deep Neural Networks

Definition 5.2.1. Let the mean-squared error be the common loss function used in deep neural networks given by;

$$Loss(x, \hat{x}) = \frac{1}{n} \sum_{i=1}^{n} (x_i - \hat{x})^2$$

with x_i being the actual value and $\hat{x}$ as the value predicted from the deep neural network after having n the number of observations.

With one input layer having two nodes and a hidden layer of three nodes, its output node will have a single node. After training data, the imputed data in the input layer will have nodes within the hidden layer.

Lemma 5.2.1. *Let $\phi(z)$ be the sigmoid activation function with its result from the calculation that passes onto different sets of layer of nodes being*

$$\phi(z) = \sigma(w_{1i}x + b_{1i}) = \left(\frac{1}{1 + e^{-(w_{1i}x+b_{1i})}} \right) \tag{5.2.1}$$

where the corresponding values from its input layer before multiplying it a weight w_j and adding a corresponding bias of b_{i1}, with 1 representing the layer and $i \in \{1,2\}$ representing its corresponding node.

Proof. From the partial derivative of the loss with respect to the individual weights; which, is split in three separate parts thus simplifying into;

$$\frac{\partial Loss(x,\hat{x})}{\partial w_i} = \left\{ \frac{\partial Loss(x,\hat{x})}{\partial \hat{x}} \right\} \left(\frac{\partial \hat{x}}{\partial y} \right) \left(\frac{\partial y}{\partial w_i} \right) \tag{5.2.2}$$

By looking at the 1st part of the derivative of equation (5.2.2), it can be simplified as;

$$\frac{\partial Loss(x,\hat{x})}{\partial w_i} = \frac{\partial}{\partial \hat{x}}\left(\frac{1}{n} \sum_{i=1}^{n} (x_i - \hat{x})^2 \right) = \frac{2}{n} \sum_{i=1}^{n} (x - \hat{x})$$

$$= \frac{2}{n} \sum_{i=1}^{n} (x_i - \hat{x}) = \sigma(z)(1 - \sigma(z))$$

$$\frac{\partial Loss(x,\hat{x})}{\partial w_i} = \sigma(z)(1 - \sigma(z)$$

and from the second part of the equation (5.2.2), we have;

$$\frac{\partial \hat{x}}{\partial z} = \frac{\partial}{\partial z}(\sigma(z)) = \frac{\partial}{\partial z}\left(\frac{1}{1+e^{-z}} \right) = \left(\frac{1}{1+e^{-z}} \right)\left(\frac{e^{-z}}{1+e^{-z}} \right)$$

$$= \left(\frac{1}{1+e^{-z}} \right)\left(1 - \frac{1}{1+e^{-z}} \right) = \sigma(z)(1 - \sigma(z))$$

$$\frac{\partial \hat{x}}{\partial z} = \sigma(z)(1 - \sigma(z)$$

And the 3rd part of the equation (5.2.2), we obtain;

$$\frac{\partial z}{\partial w_i} = \frac{\partial z}{\partial w_i}(w_i x + b) = x$$

Finally, when combined together all the derivative parts of equation (5.2.2),

$$\frac{\partial Loss(x,\hat{x})}{\partial w_i} = \frac{2}{n} \sum_{i=1}^{n} (x_i - \hat{x})\sigma(z)(1 - \sigma(z)x)$$

The backpropagation and feed-forward processes are repeated many times until the errors are negligible.

Hence, the proof. $\square$

It is essential to note that Batching can be used, which normally combines multiple observations during training data before testing to gauge the model's accuracy during the forecasting process.

5.2.2 Backward Propagation of Errors

Definition 5.2.2. Let Backward Propagation (back-propagation) be a supervised learning algorithm of artificial neural networks via gradient descent. Provided ANN and an error function, this method can calculate the error function gradient concerning the respective weights of neural networks. ANN training involves using a given unconstrained optimization problem to minimize a function within the high dimensional space. We start by defining a loss function as:

$$B = \left(\frac{\sum\limits_{i=1}^{f} (v_i - \hat{v})^2}{2} \right) \tag{5.2.3}$$

This loss function measures the deviations of predicted values $\hat{v}$ from the observed ones v i.e. it obtains the absolute error terms between these predicted values of $\hat{v}$ as well as observed values of v. The quantity B also relies on the weights of the matrices namely $Z_1, Z_2, Z_3, ... Z_f$, which ultimately influences the values of predicted $\hat{v}$. Consequently, the aim of the method is to find the exact synaptic weight values, which minimizes the value of quantity B.

While machine learning has many algorithms applied in its application, the back-propagation is among the most commonly applied in feed-forward training ANNs. The algorithm works by comparing the predicted values versus the expected ones according to modifying the synaptic weights through back-propagating the loss function's gradient.

From figure 5.1.1, the procedure continuously alternates forward with backward propagation in these steps. First, in the forward step, the predicted values of $\hat{v}$ are calculated by fixing the respective synaptic weights and in this backward step, the adjust weights to reduce the error B of the network. It is important to note that ANN can iteratively perform both forward as well as backward propagation by modifying the weights to find the combination, which diminishes the overall loss function.

Definition 5.2.3. Let the backpropagation be algorithm that can updates all weights of Z_f in the last layer by the rule of delta as follows;

$$\Delta Z_f = -i \left(\frac{\partial B}{\partial Z'_f} \right) \tag{5.2.4}$$

where i is called the learning rate. As for other preceding layers, we differentiate using product or chain rule of differentiation from Bengio, Goodfellow, and Courville (2017).

The other weights matrix Z_{f-1} are determined as:

$$\Delta Z_{f-1} = -i \frac{\partial B}{\partial Q_{f-1}} \left(\frac{\partial Q_{f-1}}{\partial Z_{f-1}} \right) \tag{5.2.5}$$

and the process continues on to many other layers in the system.

The same idea can be looked into symbolically, just like a gradient or slope descent similar to a "climbing down a steep hill" so long as it reaches a local minimum or global limit. However, the search moves in the gradient's opposite direction at every update while the gradient and learning rate slope is determined by the movement amplitude. When you want to choose from a wide range of architecture, including the hidden layers numbers, units for every layer, and the hyper-parameter values like learning rate, epochs, and activation function remain another heuristic problem for ANN users. An initial round of the hyper-parameters tuning, especially before the testing, might be highly needed. Additional extensive descriptions of ANNs and back-propagation algorithms.

5.2.3 RNN Using a LSTM Architecture

Corollary 5.2.1. *Let us incorporate the concept of Deep Learning techniques in stochastic mortality modeling thus increasing their predictability and forecasting accuracy.*

Proof. The feed-forward ANNs represent a powerful tool for analysis, can be insufficient when effectively managing time sequences of the available data. However, the recurrent connections between nodes that have featured the RNNs allow for dynamic analysis of the given sequential data. Nevertheless, through applying the given RNN structure, we often face the massive problem of gradients disappearing and weights change, before becoming tiny fast to give no effect. Consequently, the network will gradually lose its capability of learning from the past to become operationally insufficient for the more extended data sequences analysis and thus helping in making excellent predictions. This is why we say that RNNs possess a short memory only. □
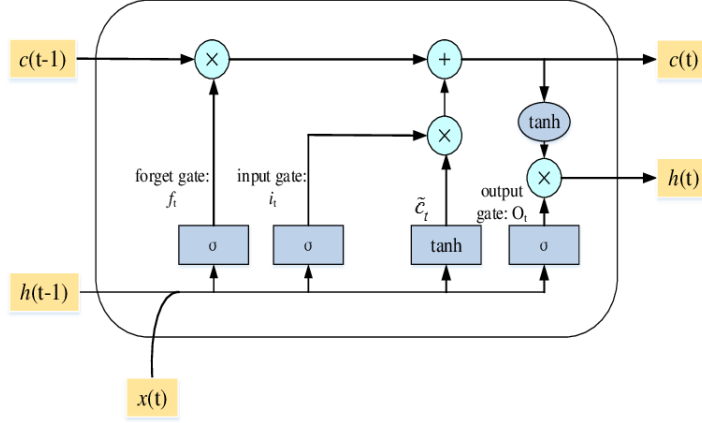
Figure 5.2.1: A LSTM Block Structure with Its Internal Information Forward Flow Design

As a way of overcoming the stated problem, (Skovajsová, 2017) had come up with LSTM. The LSTM is a version of RNN whose architecture can allow considerate relationships between the data sequence, even if it happens in long, thus eradicating vanishing gradient in the process. Similarly, RNNs need both long as well as short memory, thus efficiently generating an extraordinary performance in the time series analysis. However, several improvements in the original work, LSTM, have been improved through Lindholm and Palmborg (2022). Ultimately, one can define an excellently composed basic structure as vanilla LSTM.

Definition 5.2.4. Let $f_t = g_t$, $i_t = r_t$ and o_t be the output shown by Figure 5.2.1 that would be important in RNN analysis. Let output of the auxiliary-output gate be defined as;

$$g_t = v\left(Z_f y_t + U_f Q_{t-1} + c_f\right) \tag{5.2.6}$$

$$r_t = v\left(Z_i y_t + U_i Q_{t-1} + c_i\right) \tag{5.2.7}$$

$$o_t = v\left(Z_o y_t + U_o Q_{t-1} + c_o\right) \tag{5.2.8}$$

$$j_t = v\left(Z_j y_t + U_j Q_{t-1} + c_j\right) \tag{5.2.9}$$

The forget gate output g_t as defined by Equation (5.2.6), illustrates facts from the preceding cell state as well as the one originating from the present input are mixed within a nonlinear way through a sigmoid activation function. Afterwards, g_t is mixed through a point-wise product especially within its previous memory state $c(t-1)$. Its input gate r_t, as defined in Equation (5.2.7), uses an active sigmoid activation, which permitting for decisions when information is received before it is updated. The output gate o_t, as defined in Equation (5.2.8), plays the role of preventing non-vital memory content transmission that is stored information within the other blocks. Its role as a sigmoid function is to

pass appropriate memory information. As a way of regulating processed data flow, the input gate it does combines with that derived from all linked auxiliary NN j_t as defined in Equation (5.2.9).

Definition 5.2.5. Let the entire input block processing procedure that participates in construction of the present memory cell state denoted as:

$$c(t) = c(t-1) \circledast g_t + r_t \circledast j_t$$

To get the current output, which is a combination in between the defined function in above equation;

$$Q = \Phi(c(t)) * o_t \tag{5.2.10}$$

From equation (5.2.10), this LSTM architecture offers an outstanding tool when dealing with forecasting time series, particularly in cases of longer time lag connections, catching randomness, and management of the noise. Nevertheless, any user of LSTM, just ANNs in general, must have the face of the classical problems that concern the hyper-parameters choices.

5.3   Mathematical Application and Results

In this section, we introduced the LSTM and RNN architectures within the standard scheme of model B. More distinctly, the study's objective is to exploit the functionalities and advantages of the LSTM architecture to improve model B predictive capacity. For this aim, we design several experiments to test LSTM skills in forecasting future systematic mortality risk over time before comparing its performance with the results derived from the model of ARIMA.

Thus, the analysis of the approach will concern on the time index $k_t^{(2)}$ trend prediction, bearing in mind the ARIMA (p,d,q) model as the forecasted benchmark, whereas other parameters $k_t^{(1)}$ and $(x - \bar{x})$ are determined as per the estimation method by demonstrated by Cairns, Blake, and Dowd (2006).

Distinctly, the model applies a simple random walk process with a drift. It is key to calibrate the best ARIMA (p,d,q) as illustrated by Makridakis, Spiliotis, and Assimakopoulos (2018). In the initial round, this procedure confirms the time series stationarity using a suitable unitary root test before choosing the differencing order d. The 2nd stage chooses the auto-regressive best values and moving average order, like p and q, respectively, using information criteria for AIC or BIC.

Proposition 5.3.1. *Let $k_t^{(2)} = f(k_{t-1}^{(2)}, k_{t-2}^{(2)}, k_{t-3}^{(2)}, k_{t-4}^{(2)}, \dots k_{t-j}^{(2)}) + w_{t-x}^{(3)}$ be the performance of ARIMA (p,d,q) is compared with that of LSTM under R(y) be the ReLU activation function given by*

$$R(y) = \begin{cases} 0 & for \ y < 0 \\ y & for \ y \geq 0 \end{cases}$$

$$where \ R^{'}(y) = \begin{cases} 0 & for \ y < 0 \\ 1 & for \ y \geq 0 \end{cases} \quad with \ the \ values \ ranging \ from \ 0 \ to \ \infty \ and \ j \in N \ is$$

defined as the number of time lags being considered and $w_{t-x}^{(3)}$is the randomness term of cohort effect.

Proof. We start building an LSTM model, which enumerates the stated function f linking $k_t^{(2)}$ to the time lags. The LSTM network, just like many other common machine learning methods, needs the dataset dividing into testing and training sets. The training set often represents supervised learning, whereas testing is for validation of the model. The LSTM looks like a smooth, natural competitor to ARIMA (p,d,q) because it can capture a long-term sequence or pattern within sequential data.

Table 5.1 shows a supervised learning dataset, which is useful for prediction. Upon completion of training, the network will have learnt the input–output functional relationship thus capable of predicting future values of $k_t^{(2)}$ by using only the input. To be more practical, taking the input as $(m \times J)$ matrix with time lags of $k_t^{(2)}$ as well as the output as the $(m \times 1)$ vector of best current values, with $m \in \mathbb{N}$ is the number as in Table 5.1. $\qquad \square$

Output	Input			
$k_t^{(2)}$	$k_{t-1}^{(2)}$	$k_{t-2}^{(2)}$	.....	$k_{t-j}^{(2)}$
$k_{t+1}^{(2)}$	$k_t^{(2)}$	$k_{t-1}^{(2)}$	....	$k_{t-j+1}^{(2)}$
$k_{t+2}^{(2)}$	$k_{t+1}^{(2)}$	$k_t^{(2)}$	....	$k_{t+j-2}^{(2)}$
$k_{t+3}^{(2)}$	$k_{t+2}^{(2)}$	$k_{t+1}^{(2)}$	....	$k_{t+j-3}^{(2)}$
....	....	....	....	....
$k_{t+m}^{(2)}$	$k_{t+m-1}^{(2)}$	$k_{t+m-2}^{(2)}$	......	$k_{t+m-j}^{(2)}$

Table 5.1: Kenyan Supervised Deep Learning Dataset

The predicted $k_t^{(2)}$values, at time $m+1, m+2, m+3, ..., m+J$, are done recursively. Generally, the predicted values of $k_t^{(2)}$ in a generic time $m+t$ is determined using the values of $k_t^{(2)}$ with $t = (m+\lambda-1, m+\lambda-2, m+\lambda-3, ..., m+\lambda-J)$ as input. The values of $k_t^{(2)}$ are determined by the predicted as opposed to observed values. We start by estimating model B parameters $k_t^{(1)}$, $(x-\bar{x})$ and $k_t^{(2)}$ using the SVD method.

The extracted time series of $k_t^{(2)}$ is denoted as the first base for our analysis. The data is then split into training set and testing set as per 80% training and 20% testing rule. Consequently, we determine the last year T of observation. We have done the analysis for the U.K. and Kenya differentiating through gender with one-time lag $(j=1)$.

Nations	Number of Years	Years of Testing Set
U.K.	1930–2018	1998–2018
KENYA	**2010-2020**	**2010-2020**

Table 5.2: Testing set years as per Nation

When selecting the optimum hyperparameters combination for the neural network, it is essential to carry a preliminary fine-tuning round for all these countries while distinguishing them by gender. In this step, we can get combinations, which will be used during LSTM calibration during the forecasting procedure.

On the tuning results, we have discovered that this architecture having one hidden layer does perform better than others on our data and the number of neurons depending on the country. Using a Rectified Linear Unit (ReLU) as an activation function outperformed many other functions when testing many other countries.

Nation	Optimal ARIMA model (p,d,q)
U.K.	
Males	$ARIMA$ (1,1,0)
Females	$ARIMA$ (0,1,3)
KENYA	
Males	$ARIMA$ **(0,1,3)**
Females	$ARIMA$ **(0,1,3)**

Table 5.3: Best ARIMA by Nation and Gender

After the calibration step, the paper's analysis will include numerical and graphical processing and presentation of the goodness of fit. To be specific, the study will follow the approach of out of sample, which denotes the testing step within the field of machine learning. The estimation of parameter $k_t^{(2)}$ parameter is determined using SVD, as for male and female respectively.

Figure 5.3.1 dashed vertical line shows a separation of the forecasted period compared to one used in training the LSTM network. As for ARIMA models, it is shown that the confidence interval within 0.995 level of confidence. In addition to the graphical check, we can compare the LSTM performance against those of optimum ARIMA in the testing set before measuring the correctness of the forecasting by calculating the following measures of statistical goodness of fit; which includes MAE and RMSE:

$$MAE = \frac{\sum\limits_{\lambda=T+1}^{m-T} |k_\lambda^{(2)} - \hat{k}_\lambda^{(2)}|}{(m-T)} \tag{5.3.1}$$

71

$$RMSE = \sqrt{\frac{\sum\limits_{\lambda=T+1}^{m-T} \left(k_\lambda^{(2)} - \hat{k}_\lambda^{(2)}\right)^2}{(m-T)}} \tag{5.3.2}$$

Table 5.4 illustrates the respective performance of ARIMA and LSTM in terms of their RMSE and MAE by the individual nation and gender. From the results of measures of goodness of fit and k_t plots, we can see that the LSTM network offer excellent performances when equated to the traditional ARIMA models.

By analyzing error estimates of MAE and RMSE, Kenya shows the best performance LSTM concerning the ARIMA model for both tabulated genders. Moreover, by graphical analysis, the LSTM appears to capture the non-linearity, especially of the future mortality trends, by showing its good capability of bettering representation by decreasing mortality dynamics when dealing with the ARIMA model.

Nation		Males		Females
U.K.	MAE	RMSE	MAE	RMSE
$k_t^{(2)}$ LSTM	14.55	18.25	15.88	22.58
$k_t^{(2)}$ ARIMA	16.56	22.45	18.45	24.45
Kenya	**MAE**	**RMSE**	**MAE**	**RMSE**
$k_t^{(2)}$ **LSTM**	**18.52**	**22.35**	**20.68**	**24.07**
$k_t^{(2)}$ **ARIMA**	**20.85**	**24.05**	**21.56**	**26.86**

Table 5.4: **LSTM & ARIMA Performances in the testing set for every Nation**

Analytically, we have noticed a higher ability of the LSTM when capturing trends of non-linear without going in an opposite situation, which is an excessive oscillating or parabolic trend (as well as the latter observed when compared to traditional ANNs). Contrary, the analysis is showing that ARIMA (p, d, q) method is not effective. This evolution of $k_t^{(2)}$ as per ARIMA models is sometimes experienced out of reach within the confidence interval levels, as in the U.K. case for both sexes.

Even though ANN is an excellent and outstanding learning algorithm for modeling, it offers the only point of predictions without indicating any form of their variability. In addition, the prediction of confidence intervals is a real substantial challenge within the ANN field. Nevertheless, the LSTM network still demonstrates an excellent candidate to use when predicting the mortality trend accurately over a long time.

Table 5.4 shows that LSTM indeed over-performs the traditional ARIMA (p,d,q) model in all stated nations because of its fantastic architecture, which permits learning the vital influence from historical mortality data and replacing it with high accuracy the future years. The LSTM network capability is seen easily, particularly for the populations of the two nations where $k_t^{(2)}$ parameter does need to take a protuberant linear trend.
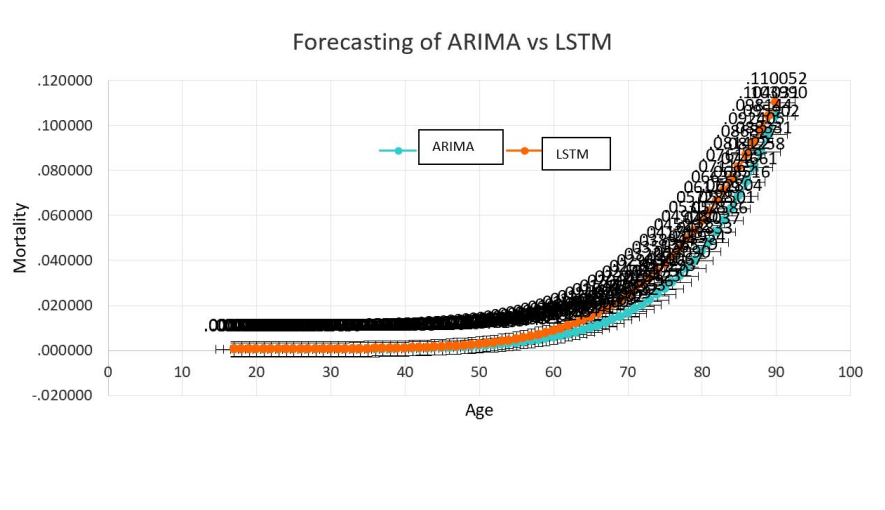
Figure 5.3.1: Mortality Prediction Under ARIMA vs LSTM

One remarkable LSTM aspect concerns the probability of achieving optimal predicting performance without resorting to a prior selection, especially of the time steps. For example, we have shown that the determined values of mortality (Figure 5.3.1) from logit-mortality rates, $logit\,\mu\,(x,t)$ for the Kenyan for males. The ARIMA model offers a trivial forecasted trend shape when compared to LSTM. The straight line of the future $k_t^{(2)}$ values, which varies over time, produces a fundamental behavior of the predicted shape of mortality.

On the contrary, the integrated model B with the LSTM has an insignificant gap between the real and the forecasted values mortality rates. From the smoothness of the curve, it is easy to prove the capability of LSTM as a better forecast on accurate and big data compared to the historical ARIMA (p,d,q) model.

5.4   Results

While the complete expectation of life has been increasing steadily globally, many theories have explained the same. However, many actuaries and demographers still don't agree on the shape and pace despite the many models used to describe it. These are some reasons why the non-linear estimation field of the time-dependent parameter in the CBD model needs to offer a crucial point, improving the model fitting by overlooking significant perspectives such as trend forecasting.

The values of MAE and RMSE from Table 5.4, under the LSTM approach, are much lower when compared to those under the traditional ARIMA models for the UK and Kenya. The accuracy of the model when forecasting the mortality models is much better in terms of precision compared to those done on the traditional ARIMA models.

73

Chapter 6

Actuarial Valuation of Life Products in Kenya

This chapter uses the calculated values of SMR incorporated Under Bühlmann's credibility approach to value the common Actuarial products such as Annuities and Assurances sold in the Kenyan market. We compare both with and those without Bühlmann's credibility incorporated approach. Results have shown that SMR under BCA has higher levels of precision compared to those without the BCA.

6.1   Life Annuities and Assurance Life Products without BCA

A prudent selection of the SMR modeling method always constitutes a vital tool in the pricing and valuation of life insurance products sold in Kenya. From the calculations from the previous chapter, we apply the obtained cohort mortality forecasts from models A, B, and C, when calculating premiums payable for the life assurance premiums.

6.1.1   Application of ATC in Kenyan Life Assurance Products

Many Kenyan life assurance and pension firms offer a wide range of life covers sold to policyholders.

Definition 6.1.1. Let us define the five most common life assurance products sold in the Kenyan insurance industry are:

1. $\bar{A}_{x+1}$ as the *EPV* of a whole life assurance of the amount of 1 that is payable instantaneously on the death's year to a life current aged $x + 1$ during his or her lifetime.

2. $A_{\underset{x+1:n]}{1}}$ as the *EPV* of a temporary endowment assurance of amount one payable at the end of the death's year to ultimate life currently aged exactly x within the period of $x+1+n$ or it death occurs in a period of n years.

3. $\bar{A}_{\underset{x+1:n]}{1}}$ as the *EPV* of a pure endowment assurance of an amount one that is payable instantaneously on survival of a ultimate life currently aged $x+1$ if that the life survives in a period of n years.

4. $\bar{A}_{x+1:n]}$ as the *EPV* of a endowment assurance of an amount one that is payable instantaneously on the death's year either on survival or death of a life currently aged $x+1$ if the life lives or dies in a period of n years.

5. $n/\bar{A}_x$ as the *EPV* of a deffered life assurance of the amount of 1 that is payable instantaneously on the death's year to a life current aged x defered for n years during his or her lifetime.

Definition 6.1.2. Let a temporary, pure and endowment assurances mathematically be defined as;

$$A_{\underset{x+1:n]}{1}} = \sum_{n=0}^{n-1} np_{x+1,t_n+1} \left(nq_{x+n,t+1+n}\right) \times (1+i)^{-(n+1)} \tag{6.1.1}$$

$$\bar{A}_{\underset{x+1:n]}{1}} = \int_{n=0}^{n-1} np_{x+1,t_n+1} (1+i)^{-n} dx \tag{6.1.2}$$

$$A_{x+1:n]} = A_{\underset{x+1:n]}{1}} + A_{\underset{x+1:n]}{1}} \tag{6.1.3}$$

where np_{x+1,t_n+1} represents the $n-year$ survival probability for a person aged x to year $t_n + 1$, while $nq_{x+n,t+1+n}$ is defined as the death probability for age x during the lifetime of a person to year $t_n + 1$, i is the applied rate of interest during the period of valuation and if $n = 0$ for the value of $np_{x+1,t_n+1} = 1$.

Each of the SMR models used, error measures for the different types of life assurance products sold in the market are given as:

$$MAE_x^{n=20} = \frac{1}{20} \sum_{x=60}^{100} |\hat{A}_{\underset{60+1:20]}{1}} - A_{\underset{60+1:20]}{1}}| \times 100 \tag{6.1.4}$$

$$MAPE_x^{n=20} = \frac{1}{20} \sum_{s=1}^{20} \sum_{x=60}^{100} |\frac{\hat{A}_{\underset{60+1:20]}{1}} - A_{\underset{60+1:20]}{1}}}{A_{\underset{60+1:20]}{1}}}| \times 100 \tag{6.1.5}$$

6.1.2 Tabulation Illustrations for the Life Assurance Products

A policyholder can buy a life assurance policy without incorporated BCA. Tables 6.1-6.5 shows the expected values of MAE as well as MAPE measures (with ranking order in the brackets) for 10-year predicted life assurance prices using the actual jump-off SMR rates for females aged 60–100 years from 2010 to 2020.

	Whole	life	Assurance
Error	A	B	C
MAE	1.986(1.982)	1.865(1.458)	1.356(1.428)
MAPE	7.235(9.128)	6.367(8.549)	7.286(8.239)

Table 6.1: EPV of MAE and MAPE measures for Males and Females (in brackets) for 10 year predicted WLA

Table 6.1 above shows the EPV of MAE and MAPE measures with (ranking order in brackets), predicted whole life assurance with actual jump-off SMR for both genders (60-100) from 2010 to 2020. In addition, model B and Model C provide better performance when compared to A when valuing whole life assurance products. It means that the lower the levels of MAE and MAPE, the better the models when valuing a whole life assurance product for both genders.

A Term	Endowment	Life	Assurance
Error	A	B	C
MAE	2.129(2.123)	1.125(1.098)	1.288(1.198)
MAPE	8.296(8.001)	6.785(16.675)	6.935(26.763)

Table 6.2: EPV of MAE and MAPE measures for Males and Females (in brackets) for 10 year predicted WLA

From Table 6.2, we deduce that model B and C offer the best estimates and forecasting for values as compared to model A for both males and females when valuing term life assurance products.

A Pure	Endowment	Life	Assurance
Error	A	B	C
MAE	1.844(2. 133)	1.218(1.328)	1.425(1.355)
MAPE	7.956(8.166)	6.825(6.866)	6.986(6.857)

Table 6.3: EPV of MAE and MAPE measures for Males and Females (in brackets) for 10 year predicted WLA

Using Table 6.3, we deduce that model B and C offer the best estimates and forecasting for values as compared to model A for both males and females when valuing pure endowment life assurance products.

An	Endowment	Life	Assurance
Error	A	B	C
MAE	1.982(2.144)	1.245(1.244)	1.256(1.276)
MAPE	7.966(38.198)	6.866(6.815)	6.988(6.914)

Table 6.4: EPV of MAE and MAPE measures for Males and Females (in brackets) for 10 year predicted WLA

From Table 6.4, we deduce that model B and E offer the best estimates and forecasting for values as compared to model A for both males and females when valuing endowment life assurance products.

A Deffered	Endowment	Life	Assurance
Error	A	B	C
MAE	2.112(32.191)	2.191(1.195)	1.195(1.266)
MAPE	8.222(8.255)	6.825(6.799)	6.252(6.988)

Table 6.5: EPV of MAE and MAPE measures for Males and Females (in brackets) for 10 year predicted WLA

Using Table 6.5, we deduce that model B and C offer the best estimates and forecasting for values as compared to model A for both males and females when valuing deffered endowment life assurance products.

6.1.3 Application of ATC in Kenyan Life Annuities Products

Since the projection of mortality models is normally used in pension mathematics applications, we use the values to calculate the performance of commonly available life annuity products.

Definition 6.1.3. Let $\ddot{a}_{x,t_n+1:n\rceil}$ be the EPV of discrete temporary life annuity-due of an amount one of an insured life aged x for a period of $t_n + 1$, which is payable yearly in advance for a period of n years so long as insured life survives during the period. The expected present value denoted as *EPV* is provided as;

$$\ddot{a}_{x,t_n+1:n\rceil} = \sum_{n=0}^{n-1} np_{x+1,t_n+1} \times (1+i)^{-n} \tag{6.1.6}$$

$$a_{x,t_n+1:n\rceil} = \sum_{n=1}^{n} np_{x+1,t_n+1} \times (1+i)^{-(n+1)} \tag{6.1.7}$$

We give an important point to note that equation (6.1.6) represents expected present value of discrete temporary life annuity-due of an amount one of an insured life aged x for a

period of $t_n + 1$, which is payable annually in arrears for a period of n years so long as insured life survives during the period.

Hence, for purposes of calculations, we apply the obtained values of estimated mortality rates from A, B, and C, after fitting from the provided data when calculating the EPV of annuities for lives aged 60–100 with $n = 10$, with an assumption of an interest rate of $i = 10$. As in our study, we apply the expected values of MAE as well as MAPE when evaluating the expected errors between predicted EPVs and those determined from the noted rates of mortality for the years 2010–2020. For every mortality models used, error measures for life annuities are given as:

$$MAE_x^{n=20} = \frac{1}{20} \sum_{x=60}^{100} |\hat{a}_{60+1:20\rceil} - a_{60+1:20\rceil}| \times 100 \qquad (6.1.8)$$

$$MAPE_x^{n=20} = \frac{1}{20} \sum_{x=60}^{100} |\frac{\hat{a}_{60+1:20\rceil} - a_{60+1:20\rceil}}{a_{60+1:20\rceil}}| \times 100 \qquad (6.1.9)$$

6.1.4 Tabulation Illustrations for the Life Annuity Products

On the illustrations on the tables, they are provided as follows;

A	Whole	Life	Annuity
Error	**A**	**B**	**C**
MAE	2.165(2.018)	1.155(1.254)	1.310(1.250)
MAPE	8.198(8.165)	6.850(6.678)	6.904(6.896)

Table 6.6: EPV of MAE and MAPE measures for Males and Females (in brackets) for a whole life annuity

A	Term	Life	Annuity
Error	**A**	**B**	**C**
MAE	2.215(2.129)	1.121(1.125)	1.219(1.267)
MAPE	8.366(8.296)	6.830(6.785)	6.958(6.920)

Table 6.7: EPV of MAE and MAPE measures for Males and Females (in brackets) for 10 year predicted temporary life annuity

Remark 6.1.1. Tables 6.6-6.7 shows that models B and A offer the best estimates and forecasting for values compared to model A for both males and females.

6.2 Life Annuities and Assurance Life Products Under BCA

6.2.1 BCA Incorporation in Kenyan Life Assurance Products

For Tabulation and illustrations, we have:

A	Whole	Life	Assurance
Error	**A**	**B**	**C**
MAE	2.255(2.132)	1.155(1.150)	1.223(1.275)
MAPE	8.290(8.268)	6.815(6.844)	6.922(6.858)
Z	0.4650	0.4665	0.4770

Table 6.8: EPV of MAE and MAPE measures for Males and Females (in brackets) for WLA

Table 6.8 deduces that models B and C offer the best estimates and forecasting for values compared to A for both males and females when valuing whole life assurance.

A Term	Endowment	life	Assurance
Error	**A**	**B**	**C**
MAE	2.005(2.135)	1.128(1.125)	1.185(1.235)
MAPE	8.198(8.388)	6.852(6.785)	6.959(6.945)
Z	0.4565	0.4670	0.4785

Table 6.9: EPV of MAE and MAPE measures for Males and Females (in brackets) for 10 year predicted temporary ELA

From Table 6.9, we deduce that models B and C offer the best estimates and forecasting for values compared to A for both males and females when valuing temporary or term endownment life assurance.

A Pure	Endowment	life	Assurance
Error	**A**	**B**	**C**
MAE	2.105(2.110)	1.198(1.125)	1.221(1.265)
MAPE	8.306(8.288)	6.800(6.765)	6.955(6.900)
Z	0.5025	0.4985	0.4765

Table 6.10: MAE and MAPE EPV measures for Males and Females (in brackets) for 10 year predicted Pure ELA

Table 6.10, we deduce that models B and C offer the best estimates and forecasting for values compared to A for both males and females when valuing pure endownment life assurance.

An	Endowment	life	Assurance
Error	**A**	**B**	**C**
MAE	2.144(2.100)	1.116(1.139)	1.275(1.269)
MAPE	8.260(8.296)	6.855(6.815)	6.900(6.955)
Z	0.4995	0.4875	0.4845

Table 6.11: MAE and MAPE EPV measures for Males and Females (in brackets) for 10 year predicted Endowment life Assurance

From Table 6.11, we deduce that models B and C offer the best estimates and forecasting for values compared to A for both males and females when valuing endownment life assurance.

A Deffered	Endowment	life	Assurance
Error	**A**	**B**	**C**
MAE	2.133(2.193)	1.145(1.118)	1.238(1.285)
MAPE	8.285(8.265)	6.766(6.756)	6.984(6.955)
Z	0.4990	0.4825	0.4970

Table 6.12: EPV of MAE and MAPE measures with (ranking order in brackets) for 10 year predicted Deffered ELA

From Table 6.12, we deduce that models B and C offer the best estimates and forecasting for values compared to model A for both males and females when valuing deffered endownment life assurance.

A life assurance product is mostly used when dealing with pension payments or those death benefits that are payable upon attaining a particular age in the future.

6.2.2 Application of BCA Incorporated in Kenyan Life Annuity Products

For tabulations and illustrations, we have;

A Whole	Life	Annuity	Product
Error	**A**	**B**	**C**
MAE	2.144(2.108)	1.249(1.146)	1.255(1.239)
MAPE	8.320(8.259)	6.820(6.726)	6.875(26.965)
Z	0.4850	0.4980	0.5010

Table 6.13: EPV of MAE and MAPE measures for Males and Females (in brackets) for a Whole life annuity

From Table 6.13, we deduce that models B and C offer the best estimates and forecasting for values compared to A for both males and females.

A Term	life	Annuity	Product
Error	A	B	C
MAE	2.144(2.136)	1.220(11.125)	1.255(1.303)
MAPE	7.965(8.305)	6.385(6.802)	6.955(6.950)
Z	0.4890	0.4855	0.4860

Table 6.14: EPV of MAE and MAPE measures for Males and Females (in brackets) for 10 year predicted temporary life annuity

From Table 6.14, we deduce that models B and C offer the best estimates and forecasting for values compared to A for both males and females. An endowment life assurance product is mostly used when dealing with pension payments or those death benefits that are payable upon attaining a particular age in the future, say $n = 10$ years.

6.3 A Comparison of Assurances and Annuities in Kenya and The UK Data

We compare the values from the Kenyan Mortality data and the UK data under BC compared with the original papers. It should help us determine how the model choice can make a difference when considering the different levels of SMR derived from the characteristics of the population such as Kenya or the UK. We note the value of Z, which is $Z = \frac{Z_K}{Z_{UK}}$ with UK=United Kingdom and K=Kenya.

6.3.1 Life assurances in Kenya vs the UK under BCA

The tables illustrate the difference in different valuations of life assurance products sold in the markets of both countries. They include the following:

A whole	life	Assurance	Product
Error	A	B	C
MAE (Kenya)	2.225(2.100)	1.150(11.205)	1.285(1.180)
MAE (UK)	2.182(2.116)	1.128(1.290)	1.299(1.125)
MAPE (Kenya)	8.356(8.296)	6.697(6.680)	6.898(6.730)
MAPE (UK)	8.452(8.296)	6.825(6.726)	6.898(6.850)
Z	0.4650	0.4665	0.4770

Table 6.15: EPV of MAE and MAPE measures for Males and Females (in brackets) for 10 year predicted WLA

Table 6.15, we deduce that models B and C offer the best estimates and forecasting for values compared to A for both males and females.

A Term	Endowment	Assurance	Product
Error	A	B	C
MAE (Kenya)	1.955(2.100)	1.018(1.102)	1.185(1.126)
MAE (UK)	2.129(2.130)	1.015(1.095)	1.196(1.132)
MAPE (Kenya)	8.128(8.126)	6.452(6.765)	6.185(6.805)
MAPE (UK)	8.093(8.225)	6.375(6.678)	6.252(6.753)
Z	0.4565	0.4670	0.4785

Table 6.16: EPV of MAE and MAPE measures for Males and Females (in brackets) for 10 year predicted Temporary ELA

Table 6.16, we deduce that models B and C offer the best estimates and forecasting for values compared to A for both males and females.

A Pure	Endowment	Assurance	Product
Error	A	B	C
MAE (Kenya)	2.013(2.115)	1.152(1.016)	1.174(1.226)
MAE (UK)	2.152(2.118)	1.143(1.018)	1.183(1.235)
MAPE (Kenya)	7.925(8.125)	6.658(6.676)	6.855(6.894)
MAPE (UK)	7.996(8.165)	6.715(6.565)	6.863(6.911)
Z	0.5025	0.4985	0.4765

Table 6.17: EPV of MAE and MAPE measures for Males and Females (in brackets) for 10 year predicted Pure ELA

Table 6.17, we deduce that models B and C offer the best estimates and forecasting for values compared to A for both males and females.

An	Endowment	Assurance	Product
Error	A	B	C
MAE (Kenya)	2.432(2.115)	1.115(1.235)	1.126(1.325)
MAE (UK)	2.315(2.125)	1.135(1.215)	1.129(1.326)
MAPE (Kenya)	8.125(8.126)	6.679(6.879)	6.851(6.785)
MAPE (UK)	8.142(8.186)	6.695(6.875)	6.900(6.820)
Z	0.4995	0.4875	0.4845

Table 6.18: EPV of MAE and MAPE measures for Males and Females (in brackets) for 10 year predicted ELA

Table 6.18, we deduce that models B and C offer the best estimates and forecasting for values compared to A for both males and females.

A Deffered	Endowment	Assurance	Product
Error	**A**	**B**	**C**
MAE (Kenya)	2.215(2.000)	1.115(1.134)	1.318(1.201)
MAE (UK)	2.225(32.095)	1.112(1.155)	1.342(1.208)
MAPE (Kenya)	8.320(8.155)	6.659(6.575)	6.895(6.785)
MAPE (UK)	8.395(8.162)	6.661(6.588)	6.935(6.800)
Z	0.4990	0.4825	0.4970

Table 6.19: EPV of MAE and MAPE measures for Males and Females (in brackets) for 10 year predicted Deffered ELA

Table 6.19, we deduce that models B and E offer the best estimates and forecasting for values compared to A for both males and females. A life assurance product is mostly used when dealing with pension payments or those death benefits that are payable upon attaining a particular age in the future.

6.3.2   Life annuities in Kenya vs the UK under BCA

For tabulations and illustrations;

A whole	life	Annuity	Product
Error	**A**	**B**	**C**
MAE (Kenya)	1.485(1.820)	1.024(1.228)	1.324(1.200)
MAE (UK)	1.505(1.919)	1.085(1.230)	1.328(1.288)
MAPE (Kenya)	6.515(8.124)	6.280(6.785)	6.350(6.925)
MAPE (UK)	6.825(8.138)	6.305(6.785)	6.695(6.898)
Z	0.4850	0.4980	0.5010

Table 6.20: EPV of MAE and MAPE measures for Males and Females (in brackets) for 10 year predicted Whole life annuity

Table 6.20, we deduce that models B and C offer the best estimates and forecasting for values compared to A for both males and females.

A Term	life	Annuity	Product
Error	**A**	**B**	**C**
MAE (Kenya)	2.025(1.995)	1.305(1.225)	1.380(1.255)
MAE (UK)	2.069(1.925)	1.312(1.250)	1.323(1.285)
MAPE (Kenya)	8.424(7.965)	6.815(6.980)	6.895(6.630)
MAPE (UK)	8.468(7.929)	6.835(6.985)	6.930(6.685)
Z	0.4890	0.4855	0.4860

Table 6.21: EPV of MAE and MAPE measures for Males and Females (in brackets) for 10 year predicted temporary life annuity

Table 6.21, we deduce that models B and C offer the best estimates and forecasting for values compared to A for both males and females.

Remark 6.3.1. Table 6.10-6.21 shows the best estimates when valuing whole-life and temporary life annuity products.

6.4 Results

The respective life assurance and annuities are calculated to show higher accuracy levels than cases when the values are without the BCA on the incorporation. From the illustration on the Kenyan data, BCA has yielded a better forecast for both males and genders (from the values of MAE and MAPE measures) compared to the A, B, and C models.

The performance is also calculated on life insurance-related products as well as applications. In addition, the applicability of the modeling approaches was comparatively illustrated on Kenyans data.

Chapter 7

Conclusions and Recommendations

7.1   Conclusions

In chapter 3, we modelled SMR Stochastically under the three-factor structure of the Age-Time-Cohort Structure for the Kenyan Population using the classical mortality models by ascertaining their respective properties under Collateral data. Our results showed that Kenyan Data do not exhibit the Normality property on the randomness used in the classical models; thus NIG proposition of modeling the disturbances of the SMR models.

The precision of the short-term forecasts or prediction outcomes was assessed by both the MAE as well as MAPE error values. All Backtesting results have shown that models B and C for males and A and B for females offer the most realistic short-term forecasts.

However, the identified parameter uncertainty also was in many cases (more evident in B for males and A for females), indicating the equivalent models' possible unsuitability, especially for long-term forecasts.

In chapter 4, we incorporated the Bühlmann Credibility Approach to Stochastic modeling of SMR while considering the NIG randomness assumptions. We have determined risk measures that show an improvement in the predictability of the models under the Bühlmann Credibility Approach, thus solving the data paucity problem.

The forecasting and pricing errors for every age and year become lower simultaneously for all the three applied mortality models, which applies to incorporating the BCA. Consequently, we conclude that the AMAPE and RMSE measures of precisions from the three unique mortality models can always be converted to a much lower and more consistent level whenever incorporating this BCA, as shown in the tables (4.7-4.9).

Furthermore, incorporation of the concept of BC into modern mortality models as a way of improving its forecasting and accuracy performance is limited to neither of the above used three mortality models nor any other common Normal assumptions, especially on the model error distributions as well as the time trend randomness or error(s).

In chapter 5, we improved the forecasting ability of the SMR models by using the Deep Learning technique. The forecasting precision of the Models is more accurate when compared to traditional statistical methods. The results should shift more researcher's attention to DL forecasting techniques compared to classical statistical methods.

We have done a deep learning CBD integrated model based on an RNN possessing LSTM architecture used to predict the future index values of the parameter, $k_t^{(2)}$. This approach has shown very high precision levels when modeling mortality trends and forecasted values instead of canonical ARIMA see in Figure (5.3.1).

In addition, LSTM has excellent features that offer more accurate forecasting while decreasing mortality trends over time than the best conventional ARIMA (p,d,q) process. Deep learning integrated systematic mortality risk modeling improves the accuracy of the models by approximately 15% for the two countries.

Ultimately, we have valued the Expected Present Value of standard Assurance and Annuities under the Integrated Bühlmann Credibility Approach for Kenyan Population.

The EPV of Assurance and Annuities under the Bühlmann Credibility Approach are more precise than those under Classical models for Kenyan data. We can conclude that, in aggregate, credibility modeling incorporated methods performed much better than the conventional models of A, B, and C, thus deciding to offer better options during forecasting the SMR.

For comparability purposes, using BCA applied to the Kenyan mortality dataset has improved the modeling and forecasting of SMR compared to the UK data. The results have played an essential role in determining modern actuarial products sold in the Kenyan market to make these products affordable for the Kenyan population, especially for low-income earning persons.

7.2 Recommendations

7.2.1 Government Policy

Government regulators like IRA and RBA, given a mandate by the Kenyan constitution to regulate pension and assurance products sold in Kenya, should institute more policies to allow insurance firms to use the Bühlmann Credibility Approach when calculating SMR in case of data paucity.

These policy adjustments will reduce the cost of these products, thus enhancing coverage among Kenyans. In addition, it will help help to reduce losses incurred from wrong estimations, thus reducing ultimate ruin probabilities for many insurance and pension firms.

The research has also shown that statistical models seem less accurate than machine learning (deep learning algorithms) during forecasting of SMR. The government agencies should adopt the application of the deep learning approach. It provides a better projection and is crucial in planning government activities for Kenyans.

Ultimately, deep learning techniques can help deal with the dynamic SMR instead of using crude mortality estimations, leading to under-estimations or over-estimations during modeling.

7.2.2   Areas of Further Research

This study has done Stochastic Modeling of Systematic Mortality Risk under Collateral Data and Its Applications. Further research can be done in the following ways;

- The research can be improved by the incorporation of the Bühlmann-Straub Credibility Model. Relaxing this means the independence of lives during modeling systematic mortality risk is not considered. This result is significant, especially when looking at the countries with similar demographic characteristics. In addition, the modeling of a four-factor SMR can play an essential role in improving the precision of the results.

- Bidirectional LSTM model where the design of BLSTM is such that it can access both the past as well as future information through the combination of both a forward hidden layer with a backward hidden layer. In addition, it can access long-range information in two opposite directions: one from bottom to top and the other processing the sequential data from top to bottom, which is likely to increase the precision of the results.

- The use of multi-level hierarchical credibility modeling using the multiple regression method when modeling mortality data of multi-population in a hierarchical Bühlmann-Straub form is another vital area of study for further study.

- With the introduction of financial futures traded in the Nairobi Securities Exchange market, Actuarial Mortality Risk management can be hedged using financial derivatives to deal with SMR. Hedging SMR under financial derivatives can form a basis for future research, especially for emerging markets like Kenya.

References

Ainou, V (2011). *A two factor Lévy model for stochastic mortality*. Tech. rep. Working Paper. http://iriaf. univ-poitiers. fr/colloque2011/article/v1s1a3. pdf.

Akaike, Hirotugu (1974). "A new look at the statistical model identification". In: *IEEE transactions on automatic control* 19.6, pp. 716–723.

Authority, Insurance Regulatory (2017). *Insurance industry annual report*.

Bengio, Yoshua, Ian Goodfellow, and Aaron Courville (2017). *Deep learning*. Vol. 1. MIT press Cambridge, MA, USA.

Blake, D and A Hunt (2016). "Basis Risk and Pension Schemes: A Relative Modelling Approach". In.

Blake, David and Andrew JG Cairns (2021a). "Longevity risk and capital markets: The 2019-20 update". In: *Insurance: Mathematics and Economics* 99, pp. 395–439.

– (2021b). "Longevity risk and capital markets: The 2019-20 update". In: *Insurance: Mathematics and Economics* 99, pp. 395–439.

Booth, Heather and Leonie Tickle (2008). "Mortality modelling and forecasting: A review of methods". In: *Annals of actuarial science* 3.1-2, pp. 3–43.

Bozikas, Apostolos and Georgios Pitselis (2020). "Incorporating crossed classification credibility into the Lee–Carter model for multi-population mortality data". In: *Insurance: Mathematics and Economics* 93, pp. 353–368.

Brouhns, Natacha, Michel Denuit, and Jeroen K Vermunt (2002). "A Poisson log-bilinear regression approach to the construction of projected lifetables". In: *Insurance: Mathematics and economics* 31.3, pp. 373–393.

Brunton, Steven L and J Nathan Kutz (2022). *Data-driven science and engineering: Machine learning, dynamical systems, and control*. Cambridge University Press.

Buhlmann, Hans and Alois Gisler (2005). *A course in credibility theory and its applications*. Springer.

Burnham, Kenneth P and David R Anderson (2004). "Multimodel inference: understanding AIC and BIC in model selection". In: *Sociological methods & research* 33.2, pp. 261–304.

Cairns, Andrew JG, David Blake, and Kevin Dowd (2006). "A two-factor model for stochastic mortality with parameter uncertainty: theory and calibration". In: *Journal of Risk and Insurance* 73.4, pp. 687–718.

Case, Anne and Angus Deaton (2017). "Mortality and morbidity in the 21st century". In: *Brookings papers on economic activity* 2017.1, pp. 397–476.

Choi, In (2001). "Unit root tests for panel data". In: *Journal of international money and Finance* 20.2, pp. 249–272.

Cramér, H and H Wold (1935). "Mortality variations in Sweden: a study in graduation and forecasting". In: *Scandinavian Actuarial Journal* 1935.3-4, pp. 161–241.

Cupido, Kyran, A Stewart Fotheringham, and Petar Jevtic (2021). "Local modelling of US mortality rates: A multiscale geographically weighted regression approach". In: *Population, Space and Place* 27.1, e2379.

Currie, Iain D (2016). "On fitting generalized linear and non-linear models of mortality". In: *Scandinavian Actuarial Journal* 2016.4, pp. 356–383.

De Moivre, A (1725). "Annuities on lives: Or, the valuation of annuities upon any number of lives as also of reversions". In: *London: William Person* 1725.

Dickey, David A and Wayne A Fuller (1979). "Distribution of the estimators for autoregressive time series with a unit root". In: *Journal of the American statistical association* 74.366a, pp. 427–431.

Doornik, Jurgen A and Henrik Hansen (2008). "An omnibus test for univariate and multivariate normality". In: *Oxford Bulletin of Economics and Statistics* 70, pp. 927–939.

Evans, Diane L, John H Drew, and Lawrence M Leemis (2017). "The distribution of the Kolmogorov–Smirnov, Cramer–von Mises, and Anderson–Darling test statistics for exponential populations with estimated parameters". In: *Computational Probability Applications*. Springer, pp. 165–190.

Gaille, Séverine (2012). "Forecasting mortality: when academia meets practice". In: *European Actuarial Journal* 2.1, pp. 49–76.

Gompertz, Benjamin (1825). "XXIV. On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies. In a letter to Francis Baily, Esq. FRS &c". In: *Philosophical transactions of the Royal Society of London* 115, pp. 513–583.

Haberman, Steven and Arthur Renshaw (2011). "A comparative study of parametric mortality projection models". In: *Insurance: Mathematics and Economics* 48.1, pp. 35–55.

Hald, Anders (1981). "TN Thiele's contributions to statistics". In: *International Statistical Review/Revue Internationale de Statistique*, pp. 1–20.

Hardy, Mary R and Harry H Panjer (1998). "A credibility approach to mortality risk". In: *ASTIN Bulletin: The Journal of the IAA* 28.2, pp. 269–283.

Hatzopoulos, P and S Haberman (2015). "Modeling trends in cohort survival probabilities". In: *Insurance: Mathematics and Economics* 64, pp. 162–179.

Hatzopoulos, P and Steven Haberman (2011). "A dynamic parameterization modeling for the age–period–cohort mortality". In: *Insurance: Mathematics and Economics* 49.2, pp. 155–174.

Hatzopoulos, Petros and Steven Haberman (2009). "A parameterized approach to modeling and forecasting mortality". In: *Insurance: Mathematics and Economics* 44.1, pp. 103–123.

– (2013). "Common mortality modeling and coherent forecasts. An empirical analysis of worldwide mortality data". In: *Insurance: Mathematics and Economics* 52.2, pp. 320–337.

Heligman, Larry and John H Pollard (1980). "The age pattern of mortality". In: *Journal of the Institute of Actuaries* 107.1, pp. 49–80.

Hunt, Andrew and David Blake (2020). "On the structure and classification of mortality models". In: *North American Actuarial Journal*, pp. 1–20.

– (2021). "On the structure and classification of mortality models". In: *North American Actuarial Journal* 25.sup1, S215–S234.

Hurvich, Clifford M and Chih-Ling Tsai (1989). "Regression and time series model selection in small samples". In: *Biometrika* 76.2, pp. 297–307.

Hyndman, Rob J and Md Shahid Ullah (2007). "Robust forecasting of mortality and fertility rates: a functional data approach". In: *Computational Statistics & Data Analysis* 51.10, pp. 4942–4956.

Jewell, William S (1975a). "The use of collateral data in credibility theory: a hierarchical model". In.

– (1975b). "The use of collateral data in credibility theory: a hierarchical model". In.

Karlis, Dimitris (2002). "An EM type algorithm for maximum likelihood estimation of the normal–inverse Gaussian distribution". In: *Statistics & probability letters* 57.1, pp. 43–52.

Kim, Joseph HT and Yongho Jeon (2013). "Credibility theory based on trimming". In: *Insurance: Mathematics and Economics* 53.1, pp. 36–47.

Klugman, Stuart A, Harry H Panjer, and Gordon E Willmot (2012). *Loss models: from data to decisions*. Vol. 715. John Wiley & Sons.

Lee, Ronald D and Lawrence R Carter (1992). "Modeling and forecasting US mortality". In: *Journal of the American statistical association* 87.419, pp. 659–671.

Levantesi, Susanna and Virginia Pizzorusso (2019). "Application of machine learning to mortality modeling and forecasting". In: *Risks* 7.1, p. 26.

Li, Johnny Siu-Hang and Mary R Hardy (2011). "Measuring basis risk in longevity hedges". In: *North American Actuarial Journal* 15.2, pp. 177–200.

Li, Johnny Siu-Hang, Rui Zhou, and Mary Hardy (2015). "A step-by-step guide to building two-population stochastic mortality models". In: *Insurance: Mathematics and Economics* 63, pp. 121–134.

Li, Nan and Ronald Lee (2005). "Coherent mortality forecasts for a group of populations: An extension of the Lee-Carter method". In: *Demography* 42.3, pp. 575–594.

Lillestol, J (2000). "Risk analysis and the NIG distribution". In: *Journal of Risk* 2, pp. 41–56.

Lindholm, Mathias and Lina Palmborg (2022). "Efficient use of data for LSTM mortality forecasting". In: *European Actuarial Journal*, pp. 1–30.

Maccheroni, Carlo and Samuel Nocito (2017). "Backtesting the Lee–Carter and the Cairns–Blake–Dowd Stochastic Mortality Models on Italian Death Rates". In: *Risks* 5.3, p. 34.

Makeham, William Matthew (1860). "On the law of mortality and the construction of annuity tables". In: *Journal of the Institute of Actuaries* 8.6, pp. 301–310.

Makridakis, Spyros, Evangelos Spiliotis, and Vassilios Assimakopoulos (2018). "Statistical and Machine Learning forecasting methods: Concerns and ways forward". In: *PloS one* 13.3, e0194889.

Mitchell, Olivia S (2020). *Building Better Retirement Systems in the Wake of the Global Pandemic*. Tech. rep. National Bureau of Economic Research.

Najafabadi, Amir T Payandeh (2010). "A new approach to the credibility formula". In: *Insurance: Mathematics and Economics* 46.2, pp. 334–338.

Odhiambo, Joab, Patrick Weke, and Philip Ngare (2021). "A Deep Learning Integrated Cairns-Blake-Dowd (CBD) Sytematic Mortality Risk Model". In: *Journal of Risk and Financial Management* 14.6, p. 259.

Opoku, Kwadwo and Minchung Hsu (2019). "Expanding Pension Coverage in Developing Countries: Incentive Scheme for Informal Workers". In.

Party, High Age Mortality Working (2015). "Continuous Mortality Investigation Limited". In.

Perks, Wilfred (1932). "On some experiments in the graduation of mortality statistics". In: *Journal of the Institute of Actuaries* 63.1, pp. 12–57.

Plat, Richard (2009). "On stochastic mortality modeling". In: *Insurance: Mathematics and Economics* 45.3, pp. 393–404.

Raalte, Alyson A van (2021). "What have we learned about mortality patterns over the past 25 years?" In: *Population Studies* 75.sup1, pp. 105–132.

Ralević, Nebojša M (2020). "Construction of c-credibility measure and application in insurance". In: *Tokovi osiguranja* 36.2, pp. 7–20.

Rutherford, Mark J, Paul C Lambert, and John R Thompson (2010). "Age–period–cohort modeling". In: *The Stata Journal* 10.4, pp. 606–627.

Salhi, Yahia and Pierre-E Thérond (2018). "Age-specific adjustment of graduated mortality". In: *ASTIN Bulletin: The Journal of the IAA* 48.2, pp. 543–569.

Salhi, Yahia, Pierre-E Thérond, and Julien Tomas (2016). "A credibility approach of the Makeham mortality law". In: *European Actuarial Journal* 6.1, pp. 61–96.

Schinzinger, Edo, Michel M Denuit, and Marcus C Christiansen (2016). "A multivariate evolutionary credibility model for mortality improvement rates". In: *Insurance: Mathematics and Economics* 69, pp. 70–81.

Shang, Han Lin, Heather Booth, and Rob J Hyndman (2011). "Point and interval forecasts of mortality rates and life expectancy: A comparison of ten principal component methods". In: *Demographic Research* 25, pp. 173–214.

Skovajsová, Lenka (2017). "Long short-term memory description and its application in text processing". In: *2017 Communication and Information Technologies (KIT)*. IEEE, pp. 1–4.

Thadewald, Thorsten and Herbert Büning (2007). "Jarque–Bera test and its competitors for testing normality–a power comparison". In: *Journal of applied statistics* 34.1, pp. 87–105.

Tsai, Cary Chi-Liang and Tzuling Lin (2017a). "A Bühlmann credibility approach to modeling mortality rates". In: *North American Actuarial Journal* 21.2, pp. 204–227.

– (2017b). "Incorporating the Bühlmann credibility into mortality models to improve forecasting performances". In: *Scandinavian Actuarial Journal* 2017.5, pp. 419–440.

Tsai, Cary Chi-Liang and Shuai Yang (2015a). "A linear regression approach to modeling mortality rates of different forms". In: *North American Actuarial Journal* 19.1, pp. 1–23.

– (2015b). "A linear regression approach to modeling mortality rates of different forms". In: *North American Actuarial Journal* 19.1, pp. 1–23.

Tsai, Cary Chi-Liang and Ying Zhang (2019). "A multi-dimensional Bühlmann credibility approach to modeling multi-population mortality rates". In: *Scandinavian Actuarial Journal* 2019.5, pp. 406–431.

Van Berkum, Frank, Katrien Antonio, and Michel Vellekoop (2016). "The impact of multiple structural changes on mortality predictions". In: *Scandinavian Actuarial Journal* 2016.7, pp. 581–603.

Villegas, Andrés, Vladimir K Kaishev, and Pietro Millossovich (2015). "StMoMo: An R package for stochastic mortality modelling". In: *7th Australasian Actuarial Education and Research Symposium*.

Wang, Chou-Wen, Hong-Chih Huang, and I-Chien Liu (2011). "A quantitative comparison of the Lee-Carter model under different types of non-Gaussian innovations". In: *The Geneva Papers on Risk and Insurance-Issues and Practice* 36.4, pp. 675–696.

Appendices

Appendix A: Kenyan Complete Life Table For Assured Mortality Lives

For this study, the new complete life table that we developed from the benchmark 2010 life table is summarized in the table below. All analyses of the research in terms of figures and tables in the thesis are developed based on these extracts for both males and females for the Kenyan population for the years 2020 and 2050. The Kenyan Male and Female population benchmark as per the year 2020 is summarized in Table 7.1 and 7.2 , respectively, as follows;

It is important to note that the force of mortality defined as $\mu(x)$ and probability of death denoted by $q(x)$ is almost similar at the young ages of Kenyan from 20 to 35 years of age, but changes sharply for the old ages from 85 to 100 years old. This means that the instantaneous death rate and probability of death are almost the same at the age of 20 to 40 but differ significantly from 85 to 100. This is attributed to many causes of death in old ages, such as less pandemic resistance like Covid-19.

These other old age-related health complications are uncommon among the younger Kenyan, as illustrated in both Males and Females as in Table 7.1 and 7.2, respectively. In addition, the force of mortality increases for both males and females as a life ages from 85 to 100 years when compared to the probability of death.

From the below tables, we have shown an Extract of the Complete life Table for Males and Females in Kenya as of 2020, as shown in Table 7.1 and Table 7.2, respectively. The Extract of the Complete life Table for Males and Females in Kenya as of 2030 is shown in Table 7.3 and Table 7.4, respectively.

Age(x)	μ(x)	q(x)	m(x)	l(x)	d(x)	L(x)	T(x)	e(x)
	Kenyan		*Female*	**Life**	**Table**			
20	0.00038	0.00038	0.00038	100,000	38	99,981	5,378,188	53.7818
21	0.00063	0.00063	0.00063	99,962	63	99,931	5,278,207	52.8021
22	0.00088	0.00089	0.00088	99,899	88	99,855	5,178,277	51.8351
23	0.00102	0.00102	0.00102	99,811	102	99,760	5,078,422	50.8804
24	0.00090	0.00090	0.00090	99,709	90	99,664	4,978,662	49.9319
25	0.00041	0.00041	0.00041	99,619	41	99,599	4,878,998	48.9766
26	0.00059	0.00060	0.00059	99,578	59	99,549	4,779,399	47.9965
27	0.00073	0.00074	0.00073	99,519	73	99,483	4,679,851	47.0247
28	0.00085	0.00084	0.00085	99,446	84	99,404	4,580,368	46.0588
29	0.00090	0.00090	0.00090	99,362	89	99,318	4,480,964	45.0974
30	0.00092	0.00092	0.00092	99,273	91	99,228	4,381,647	44.1373
31	0.00095	0.00095	0.00095	99,182	94	99,135	4,282,419	43.1774
32	0.00100	0.00100	0.00100	99,088	99	99,039	4,183,284	42.2179
33	0.00108	0.00108	0.00108	98,989	107	98,936	4,084,246	41.2596
34	0.00117	0.00117	0.00117	98,882	116	98,824	3,985,310	40.3037
35	0.00125	0.00126	0.00126	98,766	124	98,704	3,886,486	39.3504
36	0.00137	0.00137	0.00137	98,642	135	98,575	3,787,782	38.3993
37	0.00145	0.00146	0.00145	98,507	143	98,436	3,689,208	37.4512
38	0.00157	0.00156	0.00157	98,364	154	98,287	3,590,772	36.5049
39	0.00165	0.00165	0.00165	98,210	162	98,129	3,492,485	35.5614
40	0.00175	0.00174	0.00175	98,048	186	97,963	3,394,356	34.6193
...	...	...	...	...	...	...	...	...
85	0.17918	0.11744	0.17865	14,959	2454	13,732	64,361	4.3026
86	0.18876	0.13119	0.18818	12,505	2151	11,430	50,629	4.0486
87	0.20332	0.14660	0.20264	10,354	1905	9,402	39,199	3.7858
88	0.21933	0.16382	0.21852	8,449	1664	7,617	29,798	3.5267
89	0.24080	0.18301	0.23964	6,785	1452	6,059	22,181	3.2692
90	0.26390	0.20436	0.26244	5,333	1237	4,714	16,122	3.0231
91	0.28890	0.22552	0.28702	4,096	1028	3,582	11,408	2.7853
92	0.32082	0.24868	0.31792	3,068	842	2,647	7,826	2.5512
93	0.35809	0.27384	0.35573	2,226	672	1,890	5,179	2.3266
94	0.40354	0.30143	0.39829	1,554	516	1,296	3,289	2.1169
95	0.45433	0.33125	0.44623	1,038	379	848	1,993	1.9210
96	0.51184	0.36352	0.50029	659	264	527	1,145	1.7373
97	0.57620	0.39837	0.56143	395	173	309	618	1.5627
98	0.64909	0.43590	0.63065	222	106	169	309	1.3922
99	0.74625	0.47624	0.70933	116	61	85	140	1.2141
100	0.99999	0.51949	0.45455	55	25	55	55	0.9989

Table 7.1: An Extract of the new life Table for Males in Kenya in 2030

	Kenyan		Female	Life	Table			
Age(x)	μ(x)	q(x)	m(x)	l(x)	d(x)	L(x)	T(x)	e(x)
20	0.00038	0.00038	0.00038	100,000	38	99,981	5,378,188	53.7818
21	0.00063	0.00063	0.00063	99,962	63	99,931	5,278,207	52.8021
22	0.00088	0.00089	0.00088	99,899	88	99,855	5,178,277	51.8351
23	0.00102	0.00102	0.00102	99,811	102	99,760	5,078,422	50.8804
24	0.00090	0.00090	0.00090	99,709	90	99,664	4,978,662	49.9319
25	0.00041	0.00041	0.00041	99,619	41	99,599	4,878,998	48.9766
26	0.00059	0.00060	0.00059	99,578	59	99,549	4,779,399	47.9965
27	0.00073	0.00074	0.00073	99,519	73	99,483	4,679,851	47.0247
28	0.00085	0.00084	0.00085	99,446	84	99,404	4,580,368	46.0588
29	0.00090	0.00090	0.00090	99,362	89	99,318	4,480,964	45.0974
30	0.00092	0.00092	0.00092	99,273	91	99,228	4,381,647	44.1373
31	0.00095	0.00095	0.00095	99,182	94	99,135	4,282,419	43.1774
32	0.00100	0.00100	0.00100	99,088	99	99,039	4,183,284	42.2179
33	0.00108	0.00108	0.00108	98,989	107	98,936	4,084,246	41.2596
34	0.00117	0.00117	0.00117	98,882	116	98,824	3,985,310	40.3037
35	0.00125	0.00126	0.00126	98,766	124	98,704	3,886,486	39.3504
36	0.00137	0.00137	0.00137	98,642	135	98,575	3,787,782	38.3993
37	0.00145	0.00146	0.00145	98,507	143	98,436	3,689,208	37.4512
38	0.00157	0.00156	0.00157	98,364	154	98,287	3,590,772	36.5049
39	0.00165	0.00165	0.00165	98,210	162	98,129	3,492,485	35.5614
40	0.00175	0.00174	0.00175	98,048	186	97,963	3,394,356	34.6193
...	...	...	...	...	...	...	...	...
85	0.17918	0.11744	0.17865	14,959	2454	13,732	64,361	4.3026
86	0.18876	0.13119	0.18818	12,505	2151	11,430	50,629	4.0486
87	0.20332	0.14660	0.20264	10,354	1905	9,402	39,199	3.7858
88	0.21933	0.16382	0.21852	8,449	1664	7,617	29,798	3.5267
89	0.24080	0.18301	0.23964	6,785	1452	6,059	22,181	3.2692
90	0.26390	0.20436	0.26244	5,333	1237	4,714	16,122	3.0231
91	0.28890	0.22552	0.28702	4,096	1028	3,582	11,408	2.7853
92	0.32082	0.24868	0.31792	3,068	842	2,647	7,826	2.5512
93	0.35809	0.27384	0.35573	2,226	672	1,890	5,179	2.3266
94	0.40354	0.30143	0.39829	1,554	516	1,296	3,289	2.1169
95	0.45433	0.33125	0.44623	1,038	379	848	1,993	1.9210
96	0.51184	0.36352	0.50029	659	264	527	1,145	1.7373
97	0.57620	0.39837	0.56143	395	173	309	618	1.5627
98	0.64909	0.43590	0.63065	222	106	169	309	1.3922
99	0.74625	0.47624	0.70933	116	61	85	140	1.2141
100	0.99999	0.51949	0.45455	55	25	55	55	0.9989

Table 7.2: An Extract of the new life Table for Females in Kenya in 2030

	Kenyan		*Female*	Life	Table			
Age(x)	μ(x)	q(x)	m(x)	l(x)	d(x)	L(x)	T(x)	e(x)
20	0.00038	0.00038	0.00038	100,000	38	99,981	5,378,188	53.7818
21	0.00063	0.00063	0.00063	99,962	63	99,931	5,278,207	52.8021
22	0.00088	0.00089	0.00088	99,899	88	99,855	5,178,277	51.8351
23	0.00102	0.00102	0.00102	99,811	102	99,760	5,078,422	50.8804
24	0.00090	0.00090	0.00090	99,709	90	99,664	4,978,662	49.9319
25	0.00041	0.00041	0.00041	99,619	41	99,599	4,878,998	48.9766
26	0.00059	0.00060	0.00059	99,578	59	99,549	4,779,399	47.9965
27	0.00073	0.00074	0.00073	99,519	73	99,483	4,679,851	47.0247
28	0.00085	0.00084	0.00085	99,446	84	99,404	4,580,368	46.0588
29	0.00090	0.00090	0.00090	99,362	89	99,318	4,480,964	45.0974
30	0.00092	0.00092	0.00092	99,273	91	99,228	4,381,647	44.1373
31	0.00095	0.00095	0.00095	99,182	94	99,135	4,282,419	43.1774
32	0.00100	0.00100	0.00100	99,088	99	99,039	4,183,284	42.2179
33	0.00108	0.00108	0.00108	98,989	107	98,936	4,084,246	41.2596
34	0.00117	0.00117	0.00117	98,882	116	98,824	3,985,310	40.3037
35	0.00125	0.00126	0.00126	98,766	124	98,704	3,886,486	39.3504
36	0.00137	0.00137	0.00137	98,642	135	98,575	3,787,782	38.3993
37	0.00145	0.00146	0.00145	98,507	143	98,436	3,689,208	37.4512
38	0.00157	0.00156	0.00157	98,364	154	98,287	3,590,772	36.5049
39	0.00165	0.00165	0.00165	98,210	162	98,129	3,492,485	35.5614
40	0.00175	0.00174	0.00175	98,048	186	97,963	3,394,356	34.6193
...	...	...	...	...	...	...	...	...
85	0.17918	0.11744	0.17865	14,959	2454	13,732	64,361	4.3026
86	0.18876	0.13119	0.18818	12,505	2151	11,430	50,629	4.0486
87	0.20332	0.14660	0.20264	10,354	1905	9,402	39,199	3.7858
88	0.21933	0.16382	0.21852	8,449	1664	7,617	29,798	3.5267
89	0.24080	0.18301	0.23964	6,785	1452	6,059	22,181	3.2692
90	0.26390	0.20436	0.26244	5,333	1237	4,714	16,122	3.0231
91	0.28890	0.22552	0.28702	4,096	1028	3,582	11,408	2.7853
92	0.32082	0.24868	0.31792	3,068	842	2,647	7,826	2.5512
93	0.35809	0.27384	0.35573	2,226	672	1,890	5,179	2.3266
94	0.40354	0.30143	0.39829	1,554	516	1,296	3,289	2.1169
95	0.45433	0.33125	0.44623	1,038	379	848	1,993	1.9210
96	0.51184	0.36352	0.50029	659	264	527	1,145	1.7373
97	0.57620	0.39837	0.56143	395	173	309	618	1.5627
98	0.64909	0.43590	0.63065	222	106	169	309	1.3922
99	0.74625	0.47624	0.70933	116	61	85	140	1.2141
100	0.99999	0.51949	0.45455	55	25	55	55	0.9989

Table 7.3: An Extract of the new life Table for Males in Kenya in 2040

	Kenyan		Female	Life	Table			
Age(x)	μ(x)	q(x)	m(x)	l(x)	d(x)	L(x)	T(x)	e(x)
20	0.00038	0.00038	0.00038	100,000	38	99,981	5,378,188	53.7818
21	0.00063	0.00063	0.00063	99,962	63	99,931	5,278,207	52.8021
22	0.00088	0.00089	0.00088	99,899	88	99,855	5,178,277	51.8351
23	0.00102	0.00102	0.00102	99,811	102	99,760	5,078,422	50.8804
24	0.00090	0.00090	0.00090	99,709	90	99,664	4,978,662	49.9319
25	0.00041	0.00041	0.00041	99,619	41	99,599	4,878,998	48.9766
26	0.00059	0.00060	0.00059	99,578	59	99,549	4,779,399	47.9965
27	0.00073	0.00074	0.00073	99,519	73	99,483	4,679,851	47.0247
28	0.00085	0.00084	0.00085	99,446	84	99,404	4,580,368	46.0588
29	0.00090	0.00090	0.00090	99,362	89	99,318	4,480,964	45.0974
30	0.00092	0.00092	0.00092	99,273	91	99,228	4,381,647	44.1373
31	0.00095	0.00095	0.00095	99,182	94	99,135	4,282,419	43.1774
32	0.00100	0.00100	0.00100	99,088	99	99,039	4,183,284	42.2179
33	0.00108	0.00108	0.00108	98,989	107	98,936	4,084,246	41.2596
34	0.00117	0.00117	0.00117	98,882	116	98,824	3,985,310	40.3037
35	0.00125	0.00126	0.00126	98,766	124	98,704	3,886,486	39.3504
36	0.00137	0.00137	0.00137	98,642	135	98,575	3,787,782	38.3993
37	0.00145	0.00146	0.00145	98,507	143	98,436	3,689,208	37.4512
38	0.00157	0.00156	0.00157	98,364	154	98,287	3,590,772	36.5049
39	0.00165	0.00165	0.00165	98,210	162	98,129	3,492,485	35.5614
40	0.00175	0.00174	0.00175	98,048	186	97,963	3,394,356	34.6193
...	...	...	...	...	...	...	...	...
85	0.17918	0.11744	0.17865	14,959	2454	13,732	64,361	4.3026
86	0.18876	0.13119	0.18818	12,505	2151	11,430	50,629	4.0486
87	0.20332	0.14660	0.20264	10,354	1905	9,402	39,199	3.7858
88	0.21933	0.16382	0.21852	8,449	1664	7,617	29,798	3.5267
89	0.24080	0.18301	0.23964	6,785	1452	6,059	22,181	3.2692
90	0.26390	0.20436	0.26244	5,333	1237	4,714	16,122	3.0231
91	0.28890	0.22552	0.28702	4,096	1028	3,582	11,408	2.7853
92	0.32082	0.24868	0.31792	3,068	842	2,647	7,826	2.5512
93	0.35809	0.27384	0.35573	2,226	672	1,890	5,179	2.3266
94	0.40354	0.30143	0.39829	1,554	516	1,296	3,289	2.1169
95	0.45433	0.33125	0.44623	1,038	379	848	1,993	1.9210
96	0.51184	0.36352	0.50029	659	264	527	1,145	1.7373
97	0.57620	0.39837	0.56143	395	173	309	618	1.5627
98	0.64909	0.43590	0.63065	222	106	169	309	1.3922
99	0.74625	0.47624	0.70933	116	61	85	140	1.2141
100	0.99999	0.51949	0.45455	55	25	55	55	0.9989

Table 7.4: An Extract of the new life Table for Females in Kenya in 2040

Appendix B: R Codes Used in the Thesis

```
## install.packages("StMoMo")

#### Model definition ##

LC <- function(ax, bx, kt, b0x, gc, wxt, ages)

c1 <- mean(kt[1, ], na.rm = TRUE)

c2 <- sum(bx[, 1], na.rm = TRUE) list(ax = ax + c1 * bx, bx = bx / c2, kt = c2 * (kt - c1)) }

LC <- StMoMo(link = "logit", staticAgeFun = TRUE, periodAgeFun = "NP", constFun = constLC)

## Define Lee-Carter model using predefined function

LC <- lc(link = "logit")

## Define CBD model from scratch

CBD <- StMoMo(link = "logit", staticAgeFun = FALSE, periodAgeFun = c("1", f2))

## Define LR model using predefined functions

LR <- rh(link = "logit", cohortAgeFun = "1")

x <- ages t <- 1:nYears

c <- (1 - tail(ages, 1)):(nYears - ages[1])

gc <- gc - phi[1] - phi[2] * c - phi[3] * c ^ 2

kt[2, ] <- kt[2, ] + 2 * phi[3] * t

kt[1, ] <- kt[1, ] + phi[2] * t + phi[3] * (t ^ 2 - 2 * xbar * t)

ax <- ax + phi[1] - phi[2] * x + phi[3] * x ^ 2

ci <- rowMeans(kt, na.rm = TRUE)

ax <- ax + ci[1] + ci[2] * (xbar - x)

kt[1, ] <- kt[1, ] - ci[1]

## Fit LC, CBD, and LR models

KEMaleIniData <- central2initial(KEMaleData)

ages.fit <- 60:100

LCfit <- fit(LC, data = KEMaleIniData, ages.fit = ages.fit, wxt = wxt)

CBDfit <- fit(CBD, data = KEMaleIniData, ages.fit = ages.fit, wxt = wxt)

LRfit <- fit(LR, data = KEMaleIniData, ages.fit = ages.fit, wxt = wxt)

## Fit LR model using the LC parameters as starting values

LRfit <- fit(LR, data = KEMaleIniData, ages.fit = ages.fit, wxt = wxt, start.ax = LCfit$ax, start.bx = LCfit$bx, start.kt = LCfit$kt)

#### Plotting of estimated parameters ##

## Plot LC model

plot(LCfit, nCol = 3)
```

Plot CBD model

```
plot(CBDfit, parametricbx = FALSE)
```

Plot LR model

```
plot(LRfit, parametricbx = FALSE, nCol = 3)
```

Compute deviance residuals for LC, CBD and LR

```
LCres <- residuals(LCfit)
```

```
CBDres <- residuals(CBDfit)
```

Colour map of residuals for LC

```
plot(LCres, type = "colourmap", reslim = c(-3.5, 3.5))
```

Colour map of residuals for CBD

```
plot(CBDres, type = "colourmap", reslim = c(-3.5, 3.5))
```

Colour map of residuals for RH

```
LRres <- residuals(RHfit)
```

```
plot(RHres, type = "colourmap", reslim = c(-3.5, 3.5))
```

Scatter plot of residuals for LC

```
plot(LCres, type = "scatter", reslim = c(-3.5, 3.5))
```

Scatter plot of residuals for CBD

```
plot(CBDres, type = "scatter", reslim = c(-3.5, 3.5))
```

AIC and BIC for the models

```
logLik(LCfit), AIC(LCfit), BIC(LCfit)
```

CBD

```
logLik(CBDfit), AIC(CBDfit), BIC(CBDfit)
```

LR

```
logLik(APCfit), AIC(APCfit), BIC(APCfit)
```

Forecast all models

```
LCfor <- forecast(LCfit, h = 50)
```

```
CBDfor <- forecast(CBDfit, h = 50)
```

```
LRfor <- forecast(LRfit, h = 50, gc.order = c(1, 1, 0))
```

```
plot(LRfor, only.kt = TRUE)
```

Plot period index forecast of LC model

```
plot(LCfor, only.kt = TRUE) plot(LCforArima, only.kt = TRUE)
```

Plot cohort index forecast of LR model

```
plot(LRfor, only.gc = TRUE)
```

Simulation

Simulate all models set.seed(1234)

```
nsim <- 500

LCsim <- simulate(LCfit, nsim = nsim, h = 50)

CBDsim <- simulate(CBDfit, nsim = nsim, h = 50)

LRsim <- simulate(LRfit, nsim = nsim, h = 50, gc.order = c(1, 1, 0))
## Plot simulation trajectories###
par(mfrow=c(1, 3))
plot(LRfit$years, LRfit$kt[1, ], xlim = range(LRfit$years, LRsim$kt.s$years),
ylim = range(LRfit$kt, LRsim$kt.s$sim[1, , 1:20]), type = "l", xlab = "year", ylab = "kt",
main = "Period index") matlines(LRsim$kt.s$years, LRsim$kt.s$sim[1, , 1:20], type = "l", lty =
1)
## Plot cohort index
plot(LRfit$cohorts, LRfit$gc, xlim = range(LRfit$cohorts, LRsim$gc.s$cohorts),
ylim = range(LRfit$gc, LRsim$gc.s$sim[, 1:20], na.rm = TRUE), type = "l",
xlab = "year", ylab = "kt", main = "Cohort index (ARIMA(1,1,0) with drift)")
fan(t(LCsim$rates["65", , ]), start = 2020, probs = probs, n.fan = 4,
fan.col = colorRampPalette(c("black", "white")), ln = NULL)
fan(t(LCsim$rates["75", , ]), start = 2012, probs = probs, n.fan = 4,
fan.col = colorRampPalette(c("red", "white")), ln = NULL)
fan(t(LCsim$rates["85", , ]), start = 2012, probs = probs, n.fan = 4,
fan.col = colorRampPalette(c("blue", "white")), ln = NULL)
## Fanchart of q(x,t) for CBD model
matplot(CBDfit$years, t(qxt[c("65", "75", "85"), ]),
xlim = c(2010, 2030), ylim = c(0.0025, 0.2), pch = 20, col = "black",
log = "y", xlab = "year", ylab = "mortality rate (log scale)")
fan(t(CBDsim$rates["65", , ]), start = 2020, probs = probs, n.fan = 4,
fan.col = colorRampPalette(c("black", "white")), ln = NULL)
fan(t(CBDsim$rates["75", , ]), start = 2020, probs = probs, n.fan = 4,
text(2020, qxt[c("65", "75", "85"), "2020"], labels = c("x = 65", "x = 75", "x = 85"))
## Fanchart of q(x,t) for LR model
matplot(LRfit$years, t(qxt[c("65", "75", "85"), ]),
xlim = c(2010, 2020), ylim = c(0.0025, 0.2), pch = 20, col = "black",
log = "y", xlab = "year", ylab = "mortality rate (log scale)") f
an(t(LRsim$rates["65", , ]), start = 2012, probs = probs, n.fan = 4,
fan.col = colorRampPalette(c("black", "white")), ln = NULL)
text(2020, qxt[c("65", "75", "85"), "2030"], labels = c("x = 65", "x = 75", "x = 85"))
```

```r
## Use of function extractCohort

par(mfrow = c(1, 1))

plot(55:61, extractCohort(fitted(LCfit, type = "rates"), cohort = 1950),

type = "l", log = "y", xlab = "age", ylab = "q(x)", main = "Mortality rates for the 2018 cohort",
xlim = c(55,89), ylim = c(0.005, 0.12))

lines(62:89, extractCohort(LCfor$rates, cohort = 2015), lty = 2)

### Parameter uncertainty and Bootstrapping using Kenyan Mortality Data ##

## Fit LC model to KE data##

LCfit_KE <- fit(lc(), data = KEStMoMo, ages.fit = 0:89, years.fit = 2010:2020)

## Do semiparametric bootstrap of the LC model ##

LCboot_KE <- bootstrap(LCfit_KE, nBoot = 5000, type = "semiparametric")

## Plot bootstrapped parameters (Figure 12) plot(LCboot_KE, nCol = 3)

# the bootstrapped samples##

LCsimPU_KE <- simulate(LCboot_KE, h = 24)

## Forecast and simulate model without parameter uncertainty

LCfor_KE <- forecast(LCfit_KE, h = 24)

LCsim_KE <- simulate(LCfit_KE, nsim = 5000, h = 24)

## Plot prediction interval at ages 40,60 and 80 with and without ##

mxt <- KEfit_KE$Dxt / LCfit_KE$Ext

mxtHat <- fitted(LCfit_KE, type = "rates") mxtCentral <- LCfor_KE$rates

mxtPred97.5 <- apply(LCsim_KE$rates, c(1, 2), quantile, probs = 0.975)

mxtPredPU97.5 <- apply(LCsimPU_LE$rates, c(1, 2), quantile, probs = 0.975) x <- c("40", "60", "80")

matplot(LCfit_KE$years, t(mxt[x, ]),

xlim = range(LCfit_KE$years, LCfor_KE$years), ylim = range(mxtHatPU97.5[x, ],

ylab = "mortality rates (log scale)", log = "y", pch = 20, col = "black")

matlines(LCfit_KE$years, t(mxtHat[x, ]), lty = 1, col = "black")

matlines(CBDfit_KE$years, t(mxtHatPU2.5[x, ]), lty = 5, col = "red")

matlines(LRfit_KE$years, t(mxtHatPU97.5[x, ]), lty = 5, col = "red")

################################

rm(list=ls())

getwd() setwd("C:/Users/Joab/Desktop/stochastic mortality")

maledata<-read.csv(file.choose())

maledata$deathrate<-maledata$Deaths/maledata$Lives

maledata$logdeathrate<-log(maledata$Deaths/maledata$Lives)
```

summary(maledata) attach(maledata)

plot(logdeathrate~Age,type="l",main="Kenyan Male Mortality Evolution (2020)",

xlab="Age",ylab="log m(t,x)")

plot(logitqtx~Age,type="l",main="Kenyan Male Mortality Evolution (2020)",

xlab="Age",ylab="log q(t,x)") detach(maledata)

####Female

femaledata<-read.csv(file.choose()) names(femaledata)

femaledata$deathrate<-femaledata$Deaths/femaledata$Lives

femaledata$logdeathrate<-log(femaledata$Deaths/femaledata$Lives)

femaledata<-femaledata[1:41,]

summary(femaledata), attach(femaledata)

plot(logdeathrate~Age,type="l",main="Kenyan Female Mortality Evolution (2030)",

xlab="Age",ylab="log death rates")

plot(logqtx~Age,type="l",main="Kenyan Female Mortality Evolution (2030)",

xlab="Age",ylab="log q(t,x)") detach(femaledata)

plot(femaledata$logdeathrate~femaledata$Age,lty=1,type="l",col=4,

main="Kenyan Mortality Evolution (2020)",

xlab="Age",ylab="log death rates")

lines(maledata$logdeathrate~maledata$Age,lty=2,type="l",col=3,

main="Kenyan Female Mortality Evolution",

xlab="Age",ylab="log death rates") legend("topleft",c("Female","Male"),col=c(4,3),lty=c(1,2))

LCfit <- fit(lc(), data = EWMaleData, ages.fit = 55:89) #'

LCfor <- forecast(LCfit)

plot(LCfor) # #

LCfor.iarima1 <- forecast(LCfit, kt.method = "iarima", kt.order = c(1, 1 , 2)) #'

plot(LCfor.iarima1) #' #' #Lee-Carter (forecast with auto.arima) #'

LCfor.iarima2 <- forecast(LCfit, kt.method = "iarima") #' plot(LCfor.iarima2) #'

#' #CBD (Multivariate random walk with drift) #'

CBDfit <- fit(cbd(), data = central2initial(EWMaleData), ages.fit = 55:89) #'

CBDfor <- forecast(CBDfit) #' plot(CBDfor, parametricbx = FALSE) #'

#CBD (Independent Arima models) #

kt.order <- matrix(c(1, 1, 2, #ARIMA(1, 1, 2) for k[1] #' 0, 1, 1),

CBDfor.iarima <- forecast(CBDfit, kt.method = "iarima", kt.order = kt.order) #

plot(CBDfor.iarima, parametricbx = FALSE) #

#LR: Compare forecast with different models for the cohort index #'

```
wxt <- genWeightMat(55:89, EWMaleData$years, clip = 3) #'

LRfit <- fit(apc(), data = EWMaleData, ages.fit = 60:90,

LRfor1 <- forecast(LRfit) #' plot(LRfor1, parametricbx = FALSE, nCol = 3)'

LRfor2 <- forecast(LRfit, gc.order = c(0, 2, 2))

# plot(LRfor2, only.gc = TRUE) #

plot(c(LRfit$years, LRfor1$years)

###Compare Lee-Carter forecast using: #

# Fitted jump-off rates and all history for kt #

#' LCfor1 <- forecast(LCfit) #

LCfor2 <- forecast(LCfit, jumpchoice = "actual") #

LCfor3 <- forecast(LCfit, kt.lookback = 30) ##

plot(LCfit$years, (LCfit$Dxt / LCfit$Ext)["60", ],

#' xlim = range(LCfit$years, LCfor1$years), #'

ylim = range((LCfit$Dxt / LCfit$Ext)["60", ], LCfor1$rates["60", ],

#' LCfor2$rates["60", ],

LCfor3$rates["60", ]), #' type = "p", xlab = "year", ylab = "rate",

#' main = "Lee-Carter: Forecast of mortality rates at age 60") #'

LCfor1$rates["60", ], lty = 2)

# lines(LCfor2$years, LCfor2$rates["60", ], lty = 3, col = "blue")

#' legend("topright",legend = c("Fitted jump-off", "Actual jump-off",

#' "Fitted jump-off, 30 year look-back"), #' lty = 1:3, col = c("black", "blue", "red"))

#' LR forecast.fitStMoMo <-function(object, h = 50, level = c(80, 95)

#forecast kt

kt <- object$kt years <- object$years

nYears <- length(years) if (is.null(kt.lookback))

kt.lookback <- nYears if (kt.lookback <= 0) stop("kt.lookback must be positive")

kt.lookback <- min(c(kt.lookback, nYears)) yearsFor <- (years[nYears] + 1):(years[nYears] + h)

agesFor <- object$ages nAges <- length(object$ages)

years.f <- c(years[(kt.nNA+1):nYears], years.f) kt.h <-array(kt.h[, 1:kt.nNA], c(nrow(kt), kt.nNA))

#forecast

gc <- object$gc

cohorts <- object$cohorts nCohorts <- length(cohorts)

if (is.null(gc.lookback))

gc.lookback <- nCohorts if (gc.lookback <= 0) stop("gc.lookback must be positive")

gc.lookback <- min(c(gc.lookback, nCohorts))
```

```
gc.h <- gc cohorts.h <- cohorts

gc.model <- NULL gc.f <- NULL cohorts.f <- (cohorts[nCohorts] + 1):(cohorts[nCohorts] + h)

if (!is.null(object$model$cohortAgeFun))

gc.hNA <- nCohorts - gc.nNA

gc.model <- forecast::Arima(gc[(1 + nCohorts - gc.lookback):gc.nNA],

order = gc.order, include.constant = gc.include.constant)

cohorts.h <- cohorts[-((gc.nNA + 1):nCohorts)]

cohorts.f <- c(cohorts[(gc.nNA + 1):nCohorts], cohorts.f) } gc.f <- list(mean = as.vector(gc.for$mean),

lower = gc.for$lower, upper = gc.for$upper, level = level, model = gc.model, cohorts = cohorts.f)

rownames(oxt.f) <- ages

For #predict rates

rates <- predict(object, years = c(years.h, years.f), kt = cbind(kt.h, kt.f$mean),

gc = c(gc.h, gc.f$mean), oxt = cbind(object$oxt, oxt.f), type = "rates")

#Apply jump-off option forcastRates <- rates[, (nYears + 1):(nYears + h)]

fittedRates <- rates[, 1:nYears] if (jumpchoice == "actual")

{ jumpoffRates <- (object$Dxt / object$Ext)[, nYears]

forcastRates <- forcastRates * jumpoffRates / fittedRates[ , nYears] }

#' Predict method for Stochastic Mortality Models fits #

predict.fitStMoMo <- function(object, years, kt = NULL, gc = NULL, oxt = NULL, type = c("link", "rates"), ...)

{ type <- match.arg(type)

ages <- object$ages

nAges <- length(ages)

nYears <- length(years)

cohorts <- (years[1] - ages[nAges]):(years[nYears] - ages[1])

nCohorts <- length(cohorts)

#Check inputs if (object$model$N > 0){

kt <- as.matrix(kt) if (ncol(kt) == 1)

kt <- t(as.matrix(kt)) if (ncol(kt) != nYears)
```

Appendix C: Python Codes Used in the Thesis

```
Year <- line_input ( shape = c(1) , dtype = 'float40 ', name = 'Year ')

Age <- line_input ( shape = c(1) , dtype = 'int40 ', name = 'Age ')

Country <- line_input ( shape = c(1) , dtype = 'int40 ', name = 'Kenya ')

Sex <- line_input ( shape = c(1) , dtype = 'int40 ', name = 'Sex ')
```

```
Age_embed = Age %>%

line_embedding ( input_dim = 150 , output_dim = 5, input_length = 1, name = 'Age_embed ')
%>%

keras :: line_flatten () Sex_embed = Sex %>%

line_embedding ( input_dim = 2, output_dim = 5, input_length = 1, name = 'Sex_embed ') %>%

keras :: line_flatten ()

Country_embed = Country %>%

line_embedding ( input_dim = 41, output_dim = 5, input_length = 1, name = 'Country_embed ')
%>%

keras :: line_flatten () features <- line_concatenate ( list (Year , Age_embed , Sex_embed , Coun-
try_embed ))

middle = features %>%

line_dense ( units = 150 , activation = 'tanh ') %>%

line_batch_normalization () %>%

line_dropout (0.10) %>%

line_dense ( units = 150 , activation = 'tanh ') %>% line_batch_normalization () %>%

line_dropout (0.10) %>% line_dense ( units = 150 , activation = 'tanh ') %>%

line_batch_normalization () %>% line_dropout (0.10) %>%

line_dense ( units = 150 , activation = 'tanh ') %>%

line_batch_normalization () %>%

line_dropout (0.10) main_output = line_concatenate ( list ( features middle )) %>%

line_dense ( units = 150 , activation = 'tanh ') %>%

line_batch_normalization () %>%

line_dropout (0.10) %>%

line_dense ( units = 1, activation = 'sigmoid ', name = 'main_output ')

model <- keras_model ( inputs = c(Year , Age , Country , Sex ), outputs = c( main_output ))
```